

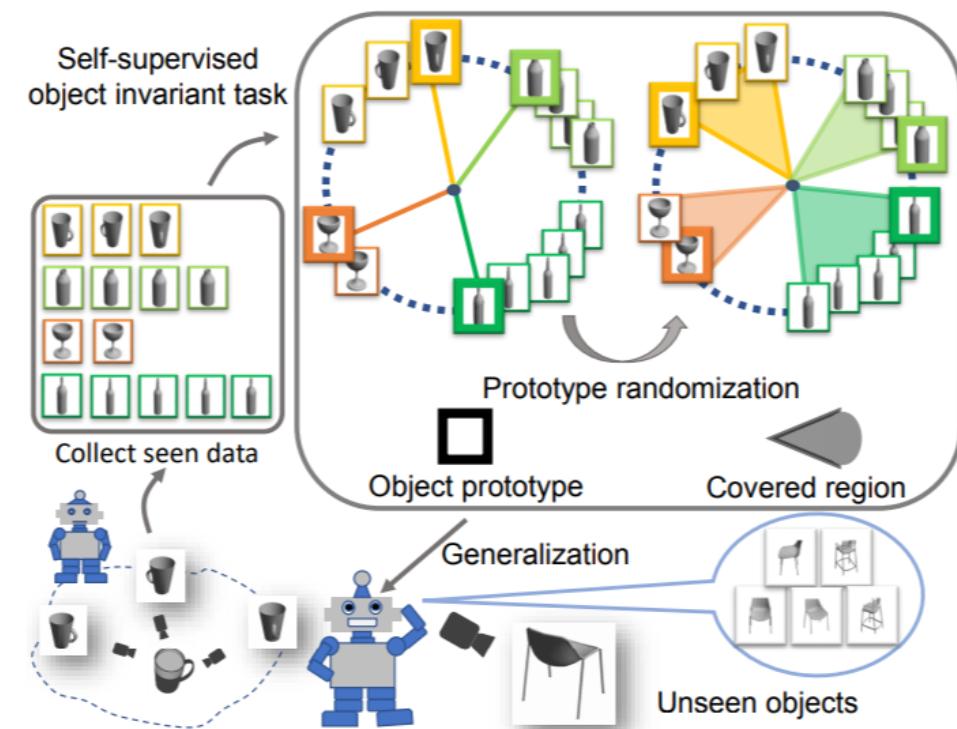
from @slam\_hub

# SLAM\_HUB Twitter 情報まとめ

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

多視点物体認識のための特徴量抽出を自己教師あり学習する手法を提案. 代理タスクとして, オブジェクトクラス分類を通し距離学習を行う. これにより視点に因らず同一オブジェクトならば埋め込み表現上でクラスターを形成. ダウンストリームタスクで他手法より高い精度を達成.

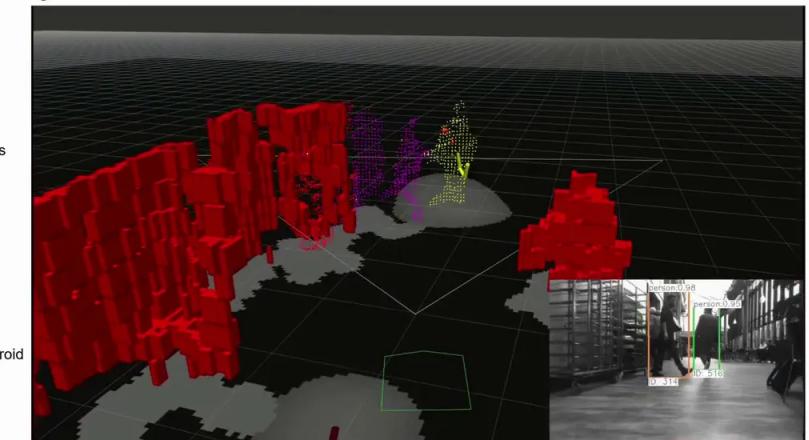
<https://arxiv.org/abs/2003.12735> pic.twitter.com/YyvMehYHKd



# Leveraging Stereo-Camera Data for Real-Time Dynamic Obstacle Detection and Tracking

ステレオカメラの情報から動的な障害を検出, 追跡するシステムの提案. ロボット周囲の物体について動的・静的の2クラスに分類. さらに動的な物体については人とそれ以外のクラスに分類する. ノイズの多いデータから高い精度の動的物体の検出, 追跡が可能なことを実験で確認. <https://youtu.be/AYjgeaQR8uQ>

Occupancy Grid



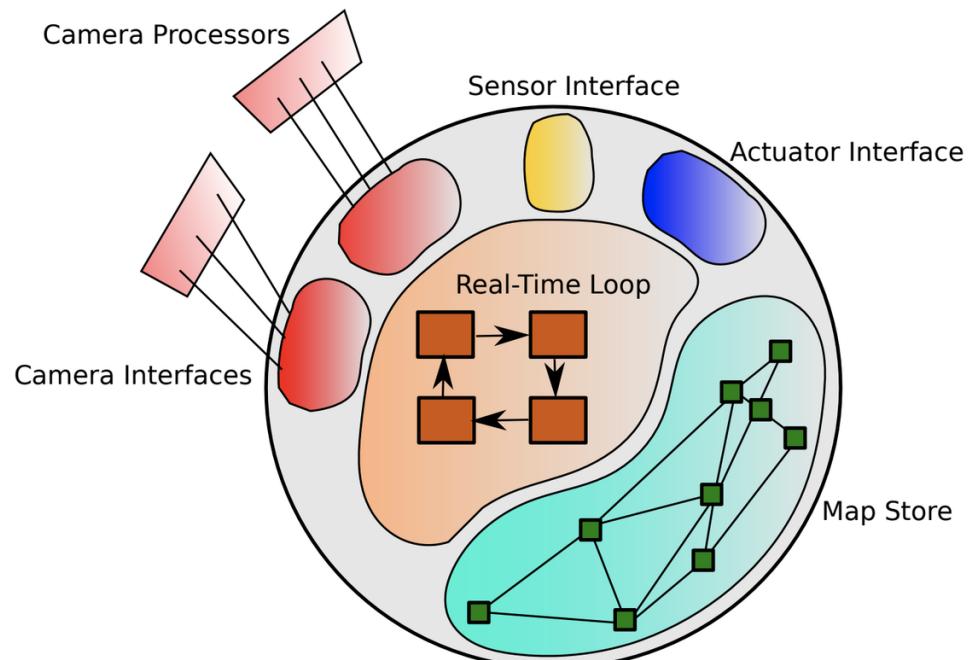
We create three maps in parallel: one for static, one for dynamic and one for uncertain objects. In each of these maps the obstacles are represented as positive costs at their respective location.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

SLAM（自己位置推定と地図構築）を発展させ、シーンやオブジェクトの関係性を理解する #SpatialAI の開発が進められている。

#SpatialAI を実際のアプリケーションに応用する上で必要なアルゴリズムやプロセッサ、センサの連携などについて提唱された最初の論文。

<https://arxiv.org/abs/1803.11288> pic.twitter.com/bm9sSOj5qN



# 3D Packing for Self-Supervised Monocular Depth Estimation [CVPR 2020]

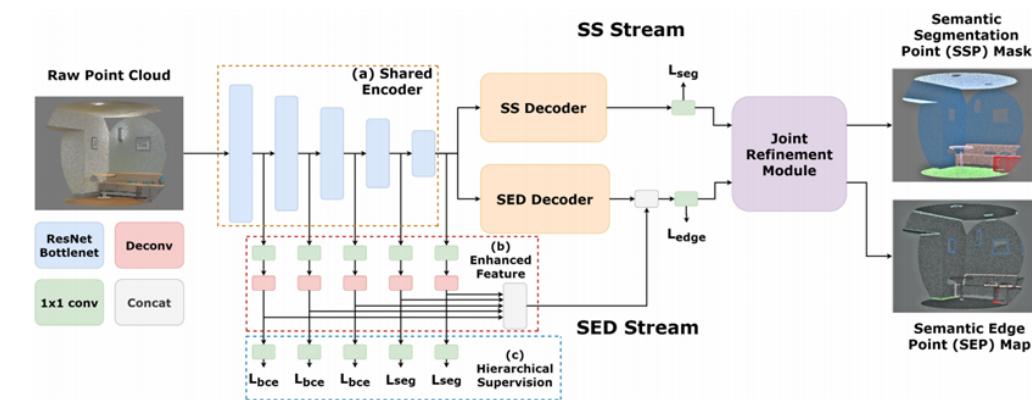
自己教師あり学習で単眼画像のデプスを推定するPackNetを提案. ピクセルをチャンネル方向に並び替えるSpace2Depthを含むPackNetにより, 重要な空間情報を保持した明瞭なデプスが推定可能. 既存の教師あり学習と同程度の精度を達成. <https://www.youtube.com/watch?v=b62iDkLgGSI> ...



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

3次元点群に対してセグメンテーションとクラス境界の推定を同時に行う手法を提案。相互に関連する両タスクを同時に解くだけでなく、双方の推定結果を用いてさらに精緻化を行うNNモジュールを提案。屋内データ (S3DIS) で SOTAを達成。

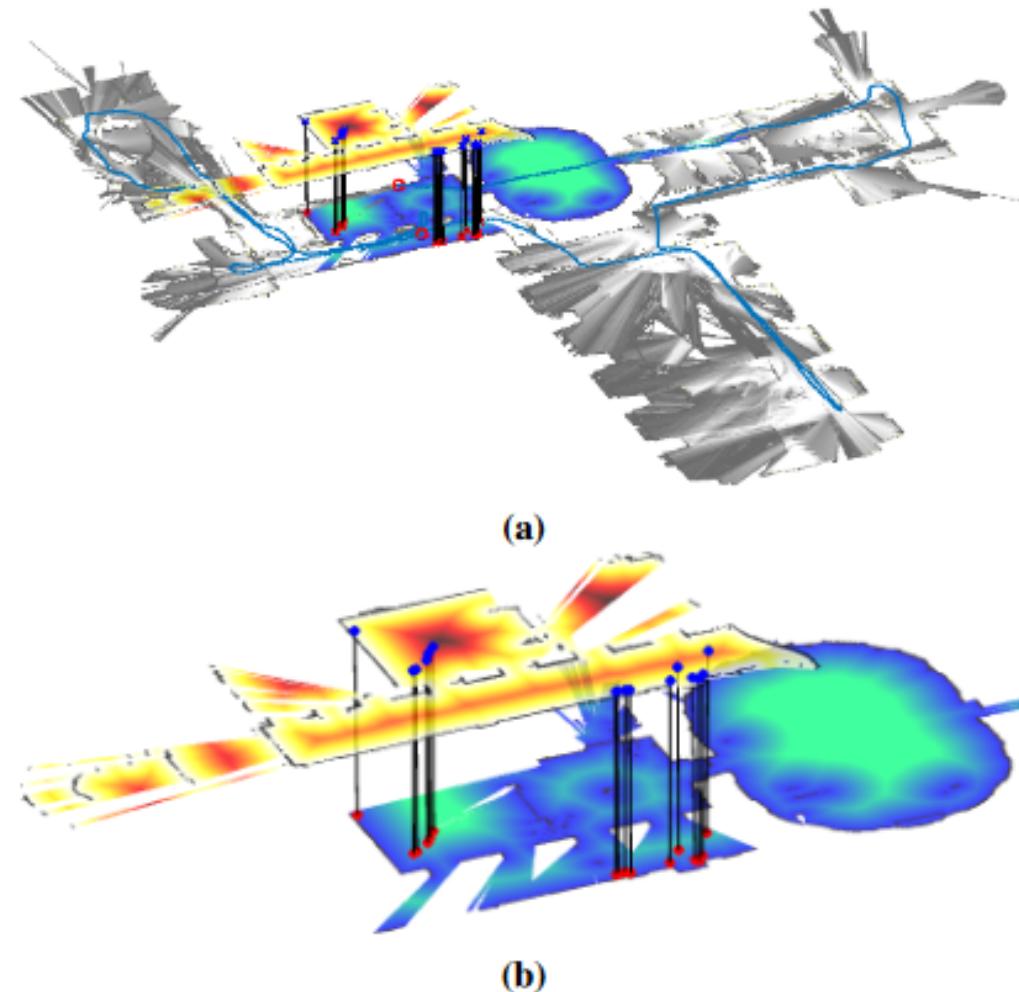
<https://arxiv.org/abs/2007.06888> pic.twitter.com/O52VDKG07v



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

2D SLAMにおける地図表現にSDFを導入し, 計測点が存在しない空間の情報を使った局所特徴(free-space features)を提案. 曲率ベースの特徴点検出と方向付き勾配ヒストグラムを使った記述子を使い, 従来手法より大域位置認識が高精度に行えることを示した.

<https://arxiv.org/abs/1908.01863> pic.twitter.com/GJSmADnz12

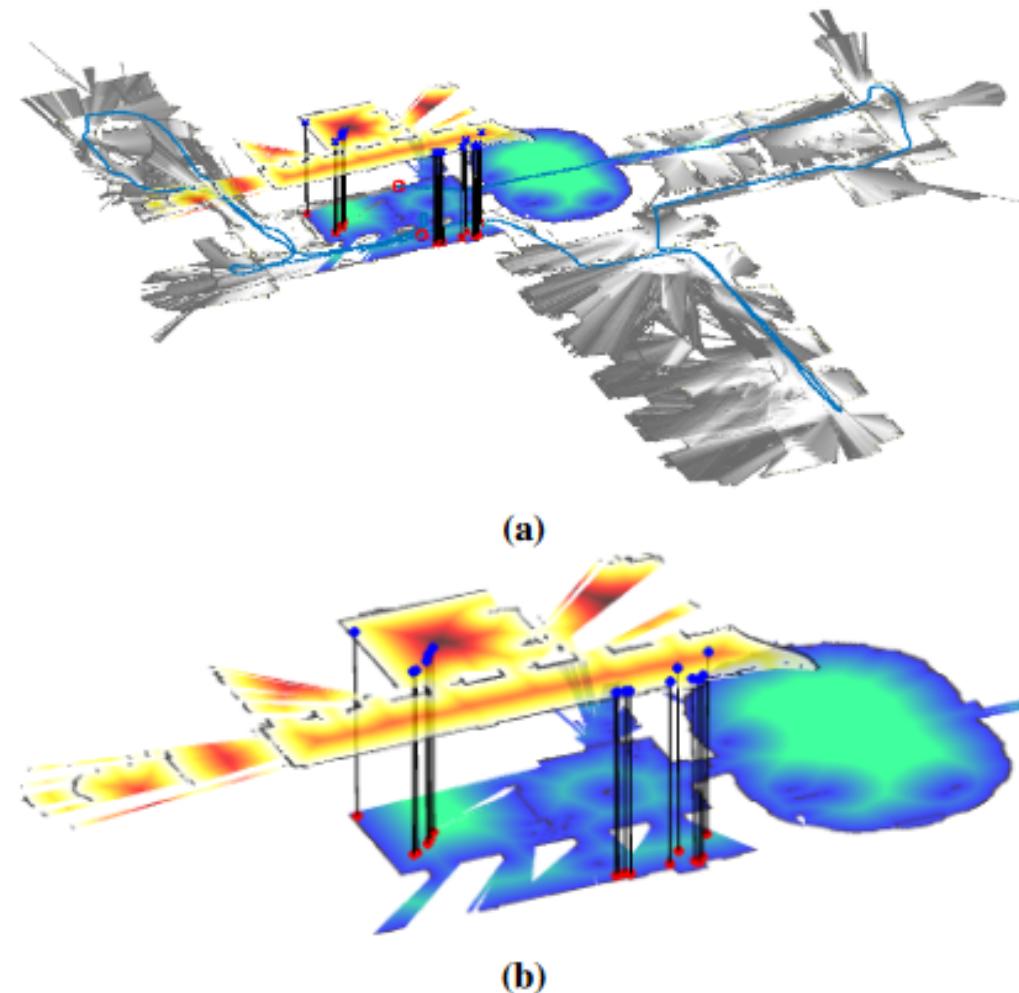


**Fig. 1:** An example of place-recognition using the proposed method. The query submap (red) from the end of a trajectory is matched against a submap (blue) from the start. Query and match submaps are displayed as distance functions, free-space feature matches are shown in black, and the path taken by the agent in blue. A close up of the matched submaps is shown in (b).

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

2D SLAMにおける地図表現にSDFを導入し, 計測点が存在しない空間の情報を使った局所特徴(free-space features)を提案. 曲率ベースの特徴点検出と方向付き勾配ヒストグラムを使った記述子を使い, 従来手法より大域位置認識が高精度に行えることを示した.

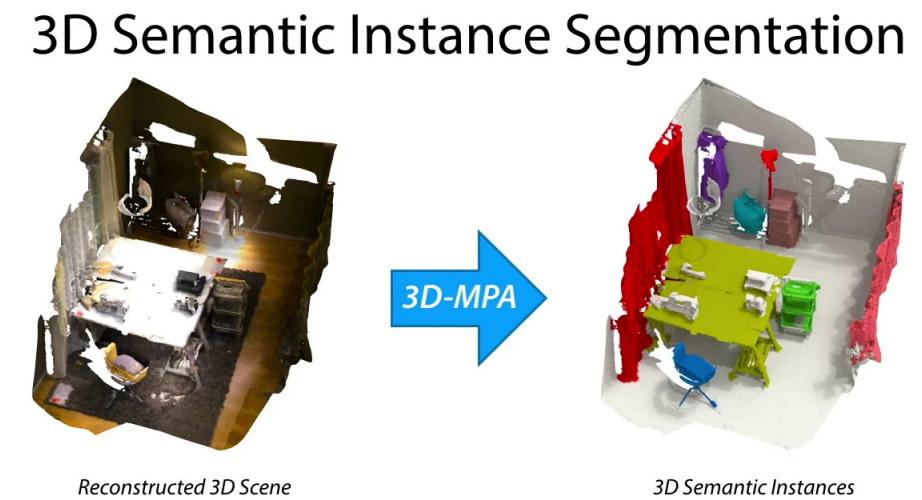
<https://arxiv.org/abs/1908.01863> pic.twitter.com/GJSmADnz12



**Fig. 1:** An example of place-recognition using the proposed method. The query submap (red) from the end of a trajectory is matched against a submap (blue) from the start. Query and match submaps are displayed as distance functions, free-space feature matches are shown in black, and the path taken by the agent in blue. A close up of the matched submaps is shown in (b).

# 3D-MPA: Multi Proposal Aggregation for 3D Semantic Instance Segmentation (CVPR 2020)

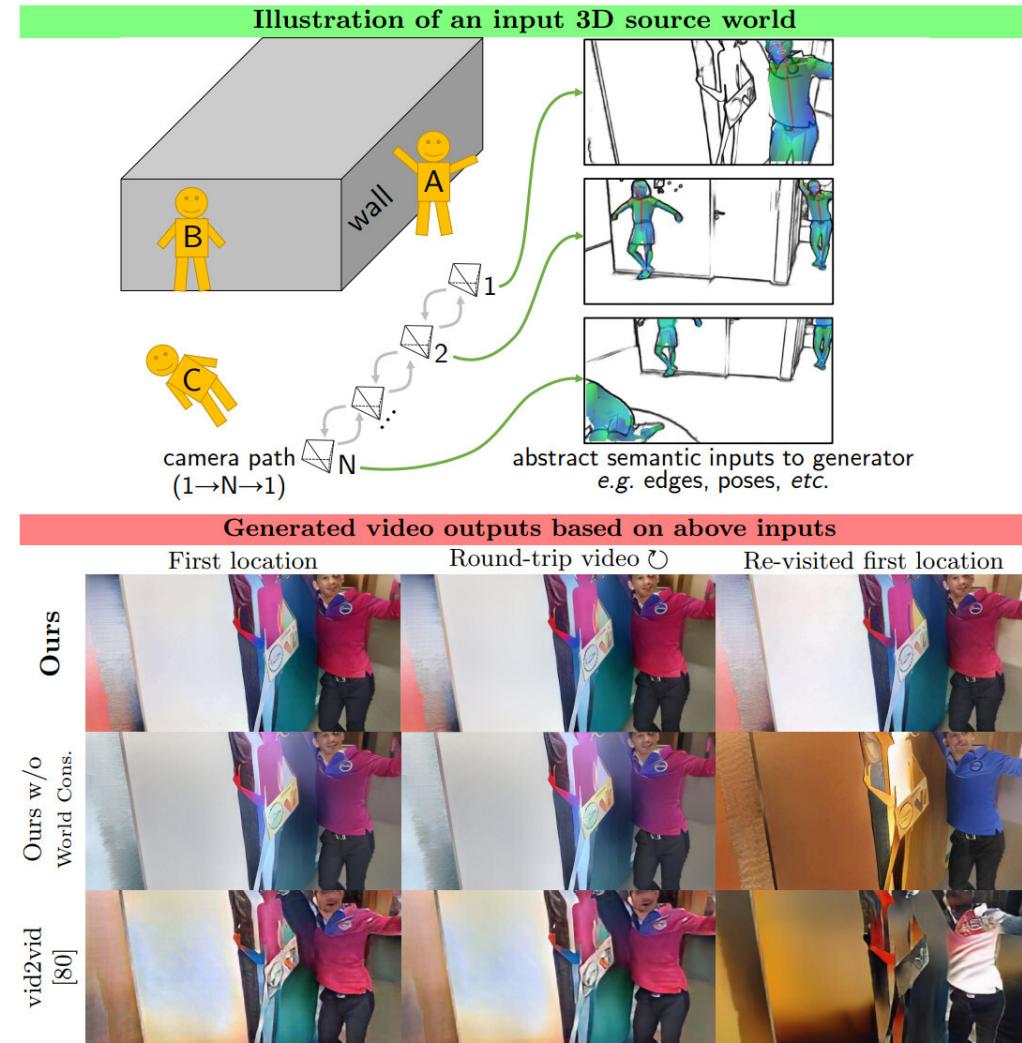
三次元点群の意味的な特徴とCenter Votesから Center Proposalを生成、GCNでProposalの特徴をリファイン、Proposalの合体によるInstance SegmentationのMulti Proposal Aggregation Network(MPA)を提案。既存手法のNon-Maximum-Suppression(NMS)と比べて、MPAの優位性を確認。<https://youtu.be/ifL8yTbRFDk>



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

大域的な一貫性を保ったvid2vid. 直前の数フレームに基づきクエリ(セマンティクス画像)に対応する画像生成を行う従来法では、同じ位置に立ち戻る場合に一貫性が保証されない。提案手法では、SfMを利用して環境を逐次的に3次元復元し、その幾何をガイドとした画像生成を行う。

<https://arxiv.org/abs/2007.08509> pic.twitter.com/uaQO8g5ofF



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Pre-Integration を用いた Tight-coupled な Visual-Inertial Odometry (VIO) に、GPS等によるグローバル座標拘束を導入する手法を初めて提案。従来法はVIOの後段にカルマンフィルタ等を用いて分割して対処。提案手法はコスト関数に拘束を組込み一括で最適化。

<https://arxiv.org/abs/2003.04159> pic.twitter.com/Im1bHSvCyO

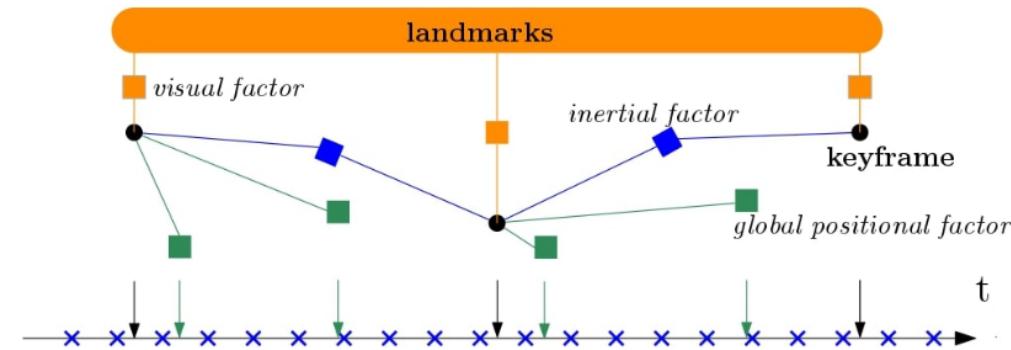
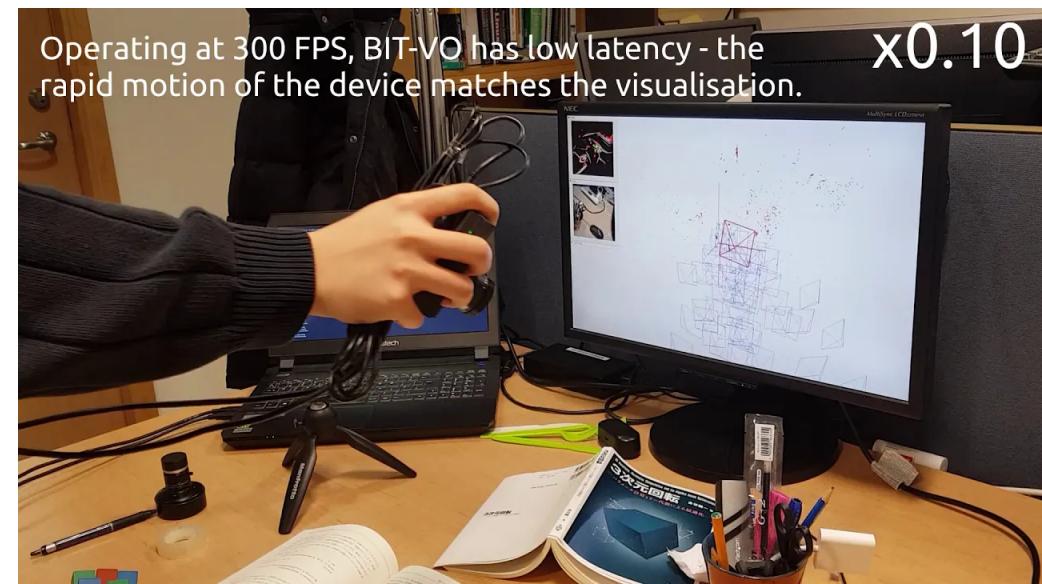


Fig. 1. Representation of the proposed optimization-based multi-sensor fusion. We distinguish three types of factors: visual (orange), inertial (blue), and global positional factors (green). The optimization variables are the states of the keyframes in the current sliding window and the visible landmarks. In the bottom part of the figure, IMU measurements are depicted with crosses on the temporal line, while keyframes and global positional measurements are depicted with black and green arrows, respectively.

# BIT-VO: Visual Odometry at 300 FPS using Binary Features from the Focal Plane

画素毎に並列計算可能なFocal-plane Sensor-processor (FPSP) を用いたVisual OdometryアルゴリズムBIT-VOを提案。FPSP上で、アナログ信号の領域で2値のエッジ検出を行い、そのエッジからバイナリ特徴を計算。それらをホストデバイスに転送することで300fpsの6DOF VOを実現。[https://www.youtube.com/watch?v=tnPfbJaPrSQ&feature=emb\\_title](https://www.youtube.com/watch?v=tnPfbJaPrSQ&feature=emb_title) ...



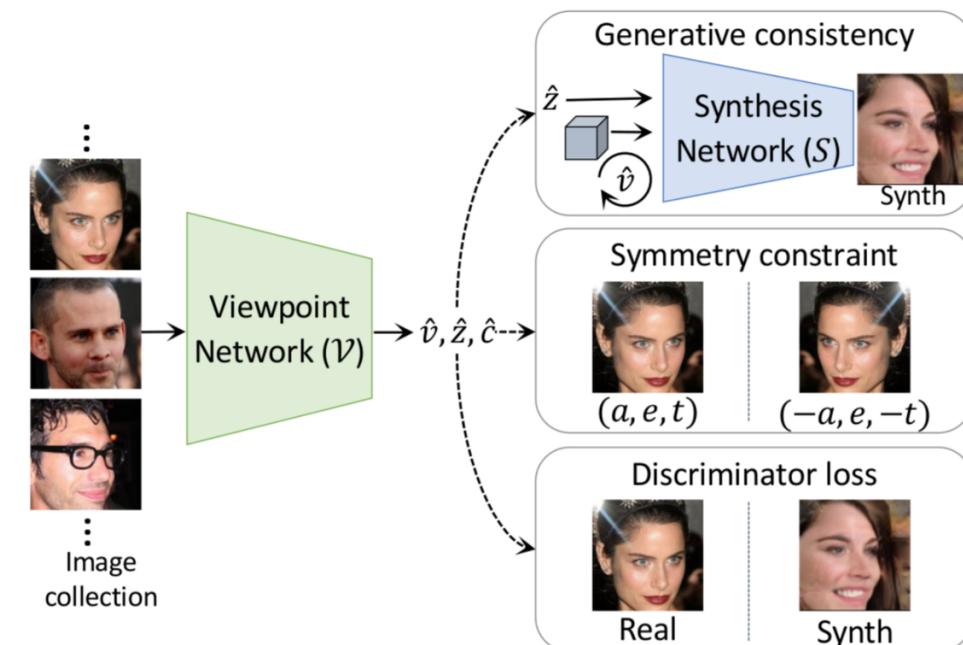
Operating at 300 FPS, BIT-VO has low latency - the rapid motion of the device matches the visualisation.

x0.10

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

画像からの物体方向推定を, self-supervisedで学習する枠組みを提案. 画像から3次元方向とスタイル特徴量を抽出し, それらの潜在変数を元に幾何学的変換を行うGenerator(GAN)を用いて学習. 損失には一貫性と水平対称性を利用し, 教師あり学習に匹敵する性能を達成.

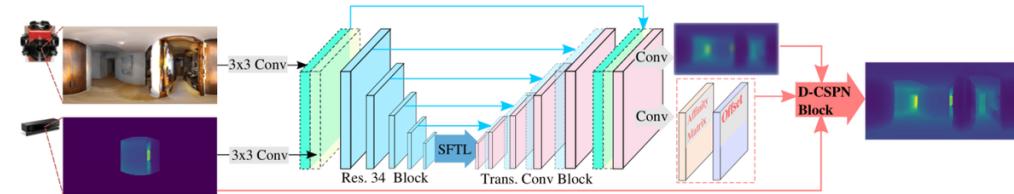
<https://arxiv.org/abs/2004.01793> pic.twitter.com/Qr4LM3uTxA



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Kinect等のPerspectiveなdepthセンサーと全方位画像を用いて全方位のdepthを得る手法の提案. Encoderの最後の層でPerspective座標に変換し特徴量の学習難度を下げる、 Decoderで equirectangular座標に戻す.他のSoTAな手法より優れていることを示した.

<https://arxiv.org/abs/2007.01475> pic.twitter.com/ZYPB7ysvaY



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

360度屋内画像における物体検出とクラス認識に関するデータセットを提示. Equirectangular形式における極領域の歪みに対応するため, Sphere Netをはじめとする球状CNNを用いたモデルで評価したところ, 透視投影画像によるデータセットでの学習よりも大きな改善が見られた.

<https://arxiv.org/abs/1910.01712> pic.twitter.com/buiy0FHupl

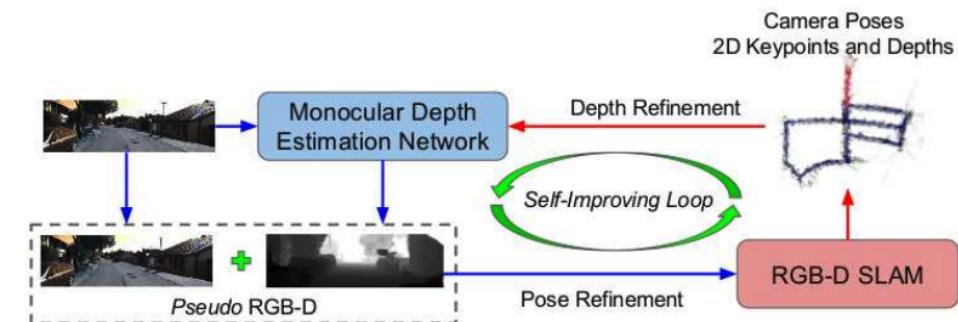


Figure 1: Examples of the images overlaid with labeled Bounding Field-of-Views (BFoVs) as well as object categories (text and color-code) in our 360-Indoor dataset.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

CNNで推定した擬似的なデプスマップを用いてRGB-D SLAMを行う。単眼デプス推定の欠点であるスケールの不整合性を、特徴点ベースのSLAMで作成された三次元点を用いてリファインする。これにより、両者の欠点を補った高精度な姿勢推定が可能となった。

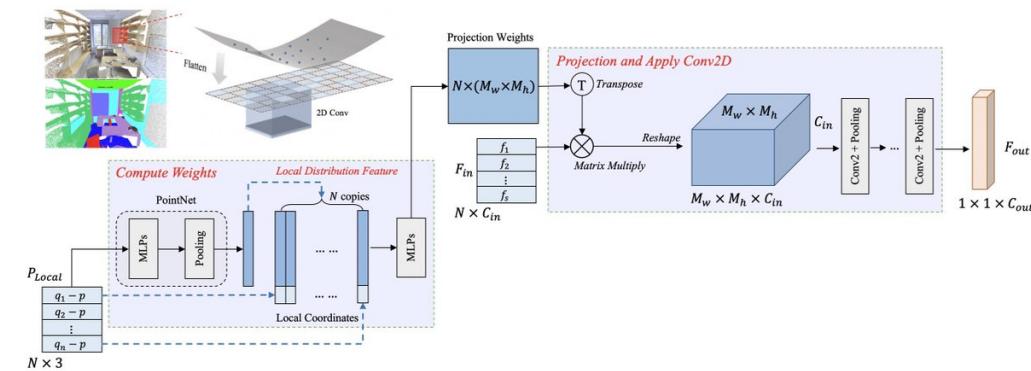
<https://arxiv.org/abs/2004.10681> pic.twitter.com/OXOuG5onDG



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

平面への投影を用いた点群畳み込みを提案. 明示的に接平面を推定するTangentConvとは異なり, 点群の投影と内挿を单一の重み行列で表現し, MLPを用いて学習ベースで推定する. Volmetricな畳み込みとの組み合わせでSoTA達成.

<https://arxiv.org/abs/2002.10701> pic.twitter.com/ZmJ3Uuwlg1



# 1-Day Learning, 1-Year Localization: LiDAR localization using Scan Context Image (RA-L + ICRA 2019)

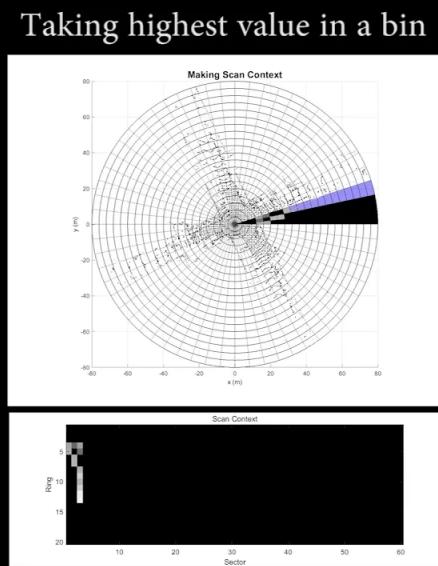
LiDARを使った位置推定手法. 極座標で作られた高さマップ(Scan Context)を入力とし, CNNで地図上での位置をクラスとして推定する. 複数の実データセット上で, 一日の学習データで一年通した長期位置推定が高精度に可能であることを示した.

<https://youtu.be/apmmduXTnaE>

Scan Context Image (SCI)  
Generation

Step 1:  
Point cloud  
to Scan Context  
[IROS 18]

Giseop Kim and Ayoung Kim, Scan Context:  
Egocentric Spatial Descriptor for Place Recognition  
within 3D Point Cloud Map, IROS, 2018



# BSP-Net: Generating Compact Meshes via Binary Space Partitioning, CVPR 2020, oral presentation video

教師なしでコンパクトかつウォータータイトな三次元メッシュ生成手法を提案。Binary Space Partitioning (BSP) で再帰的に入力形状を超平面に分解し、Constructive Solid Geometry (CSG) のブーリアン演算で、分解した超平面から複雑な表面やオブジェクトの生成が可能となる。

<https://youtu.be/9-ixexpjN-8>



BSP-Net: Generating Compact Meshes  
via Binary Space Partitioning

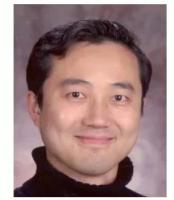
Zhiqin Chen  
Simon Fraser University



Andrea Tagliasacchi  
Google Research

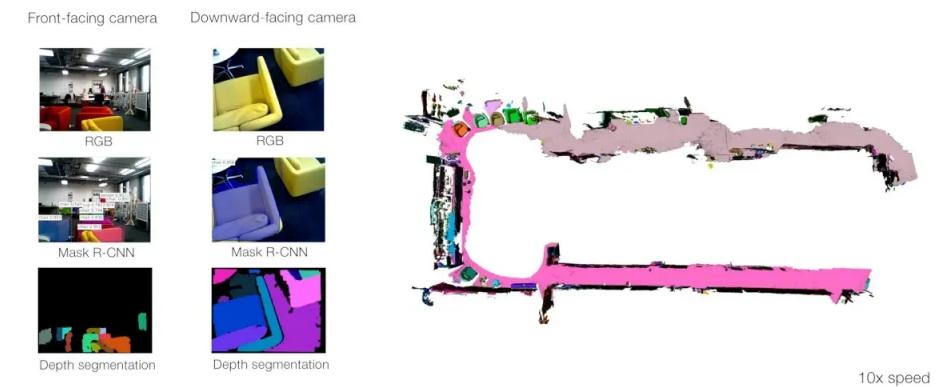


Hao (Richard) Zhang  
Simon Fraser University



# Volumetric Instance-Aware Semantic Mapping and 3D Object Discovery

インスタンスを意識したRGB-Dセンサによる地図生成. 幾何的なセグメンテーション結果をMask R-CNNから補正し, Over-segmentationを抑制した個別の物体形状を獲得. 各物体の幾何を大域地図上で関連付けていくことで, セマンティクス&インスタンス情報も付与した地図を生成. <https://youtu.be/Jvl42VJmYxg>



Reconstructing the scene geometry with a voxel size of 2 cm, the mapping framework processes incoming RGB-D image pairs at a speed of approximately 1 frame per second.

# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Visual-Inertial SLAM の Basalt。IMU の preintegration ではなく、非線形因子復元を行って大域的に最適化。IMU の積分は誤差が大きい問題に対処。VIO の相対位置拘束とローラル・ピッチ拘束、バンドル調整のループ拘束を統合。小さな最適化問題として定式化でき、精度も向上。

<https://arxiv.org/abs/1904.06504> pic.twitter.com/mQYfoQ4yjd

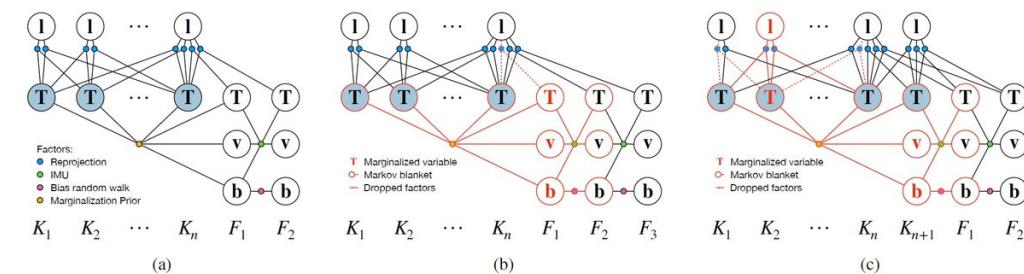


Fig. 5: Factor graphs. (a) After marginalizing a frame, the system consists of  $n$  older keyframes  $K_1 \dots K_n$  and the  $m - 1$  most recent frames  $F_1$  and  $F_2$  (which could potentially also host landmarks and hence be keyframes). After a new frame has been added, the oldest velocity  $v$  and the oldest bias  $b$  are marginalized. If they do not belong to a keyframe (b), the whole frame including its pose  $T$  is marginalized. If they belong to a keyframe (c), another keyframe is selected for marginalization, including the landmarks hosted in it and its pose. In both cases, reprojection factors where the target frame is the marginalized frame are dropped. In the latter case, reprojection factors from the marginalized frame to  $F_2$  are dropped to allow relinearization. Note that not all possible combinations of host and target frames for reprojection factors are shown.

## 9099049

任意の環境形状を表せる陰関数表現を用い2D SLAMを定式化. 陰関数に対する分散の導出や陰関数境界内外での最適化の安定化などを行い, 楕円・直線モデルを用いた評価実験では従来のモデルフィッティングベース手法より良い精度を示した.

<https://ieeexplore.ieee.org/document/9099049> ...pic.twitter.com/HnpTxH7bL6

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

複数マップ間のループ検出における外れ値検出手法を提案. 各ノード間の回転から閉ループの幾何的整合性をチェックすることで確率的に外れ値を検出. さらにEMアルゴリズムを用いてパラメータをfine-tune. 確率伝搬法より高い精度を達成し, 収束性も保証.

<https://arxiv.org/abs/2002.02638> pic.twitter.com/zugcqk2JzE

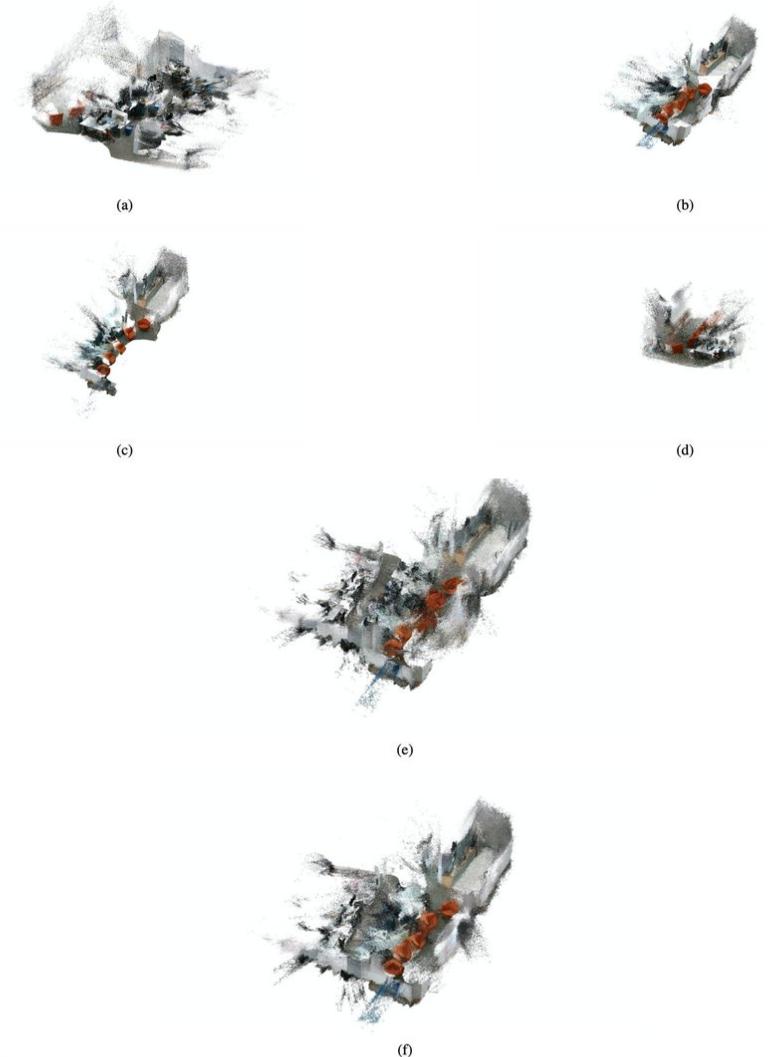


Fig. 4: In (a)-(d), four pointclouds from four different sequence of images are obtained. Relative pose measurements between a subset of images are obtained through feature matching and the use of an object detector for place recognition. After removing the outliers and joining the maps, an optimization problem is solved to reduce the overall error. The final pointcloud made from joining all the four pointclouds without outlier detection is shown in (e) and with outlier detection is shown in (f).

# Vid2Curve: Simultaneous Camera Motion Estimation and Thin Structure Reconstruction (SIGGRAPH 2020)

ワイヤーフレームや電線など細い物体に対する形状復元の手法を提案。視点の追加ごとに、点群表現のカーブとカメラポーズを新しいマッチング手法に基づき交互に最適化。また、オクルージョンを検知し誤対応を防止。カーブの各部で太さを推定することで高品質な復元が可能に。[https://youtu.be/dl2FZG\\_txN0](https://youtu.be/dl2FZG_txN0)

Cat



Reference image

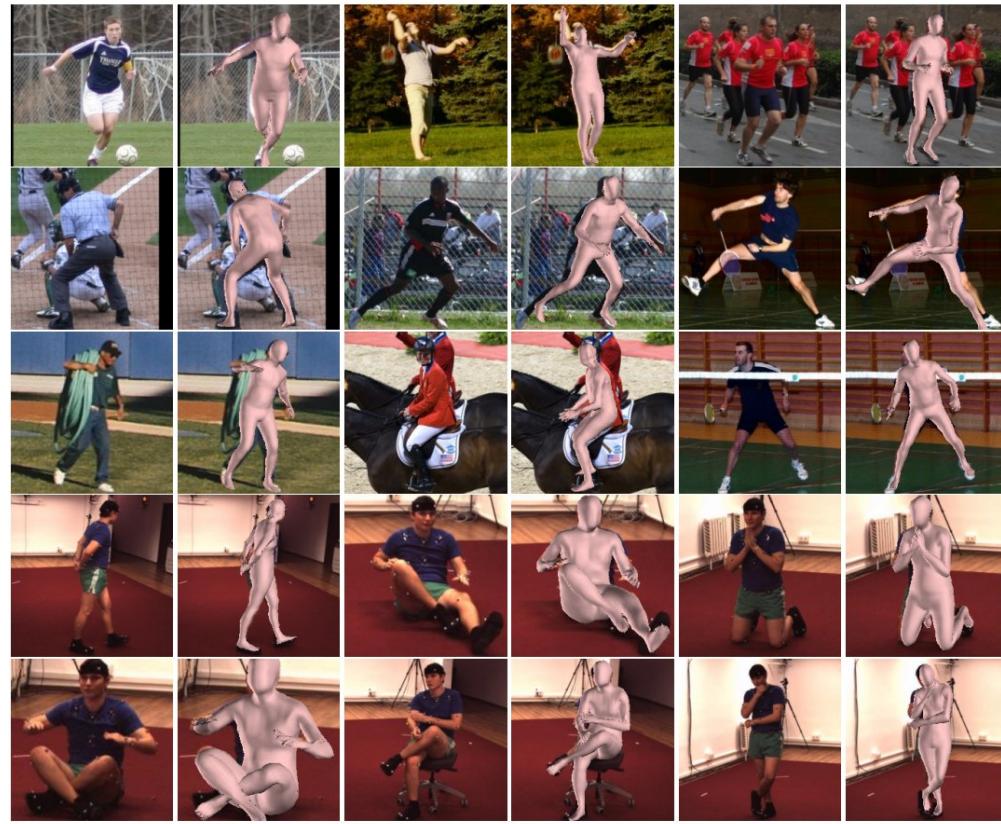


Our result

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

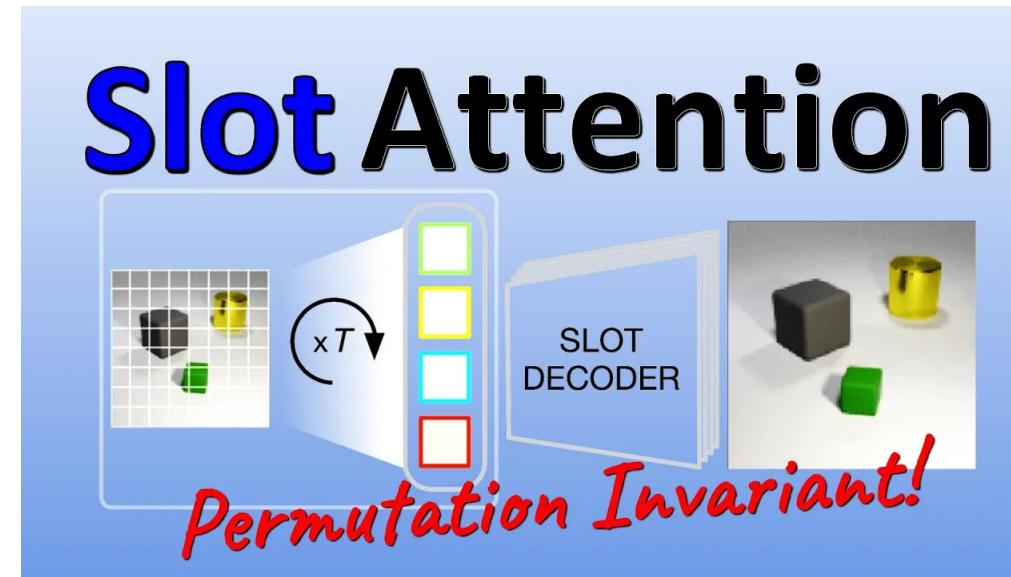
一枚の画像から人体の3D Meshを推定する手法.画像ピクセルと表面間の密な対応を推定し,その対応により画像空間からUV空間へ局所的な特徴が移され,位置マップに回帰される.最後にマッピング関数により3D Meshを再構成する.3D Meshベースの従来手法より優れていることを示した.

<https://arxiv.org/abs/2006.05734> pic.twitter.com/Ku422DMyAp



## Object-Centric Learning with Slot Attention (Paper Explained)

シーン分解や集合予測のアーキテクチャに統合可能な，オブジェクト中心の抽象表現を学習するSlot Attentionモジュールを提案. CNNの出力と構造表現間において，順列不变なk個のSlotを生成. 反復的注意メカニズムでSlotのグループ化戦略を学習. 点群やグラフのグループ化も可能 <https://youtu.be/DYBmD88vpiA>

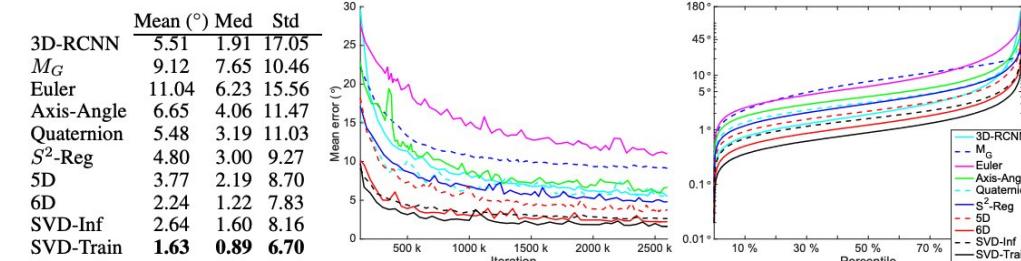


# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

深層学習における性質の良い回転表現を提案。  
回転行列を一度9パラメータで表現し、SVDによる特殊直交化により $SO(3)$ 空間へマップする。

深層学習タスクにおいてクオータニオンやangle-axisベクトルなどの他の回転表現より高精度に姿勢を求めることが可能。

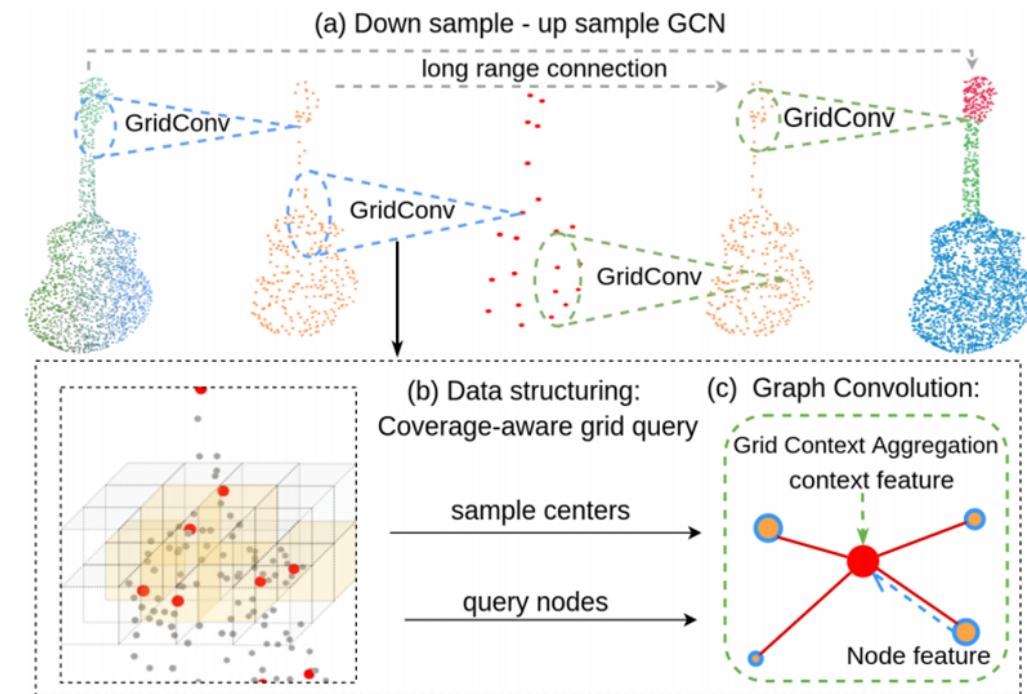
<https://arxiv.org/abs/2006.14616> pic.twitter.com/WVhQcNGBad



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

高速でスケーラブルな点群データの処理機構を提案. Voxelを用いることで高速かつカバー率の高いサンプリングを行い, Voxel内部でローカルにグラフを構築して畳み込む. 点群の分類とセグメンテーションで従来手法より高速かつ高精度を達成.

<https://arxiv.org/abs/1912.02984> pic.twitter.com/nllsA1sC2v



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

PointNetに類似したネットワークを用い点群から三角形メッシュを生成。入力点群から三角形群を出力するネットと、入力三角形群の中から3Dモデルとして妥当な三角形を判定するネットを交互に適用し、メッシュモデルを復元する。

<https://arxiv.org/abs/2005.02138> pic.twitter.com/8JeOljgxXg

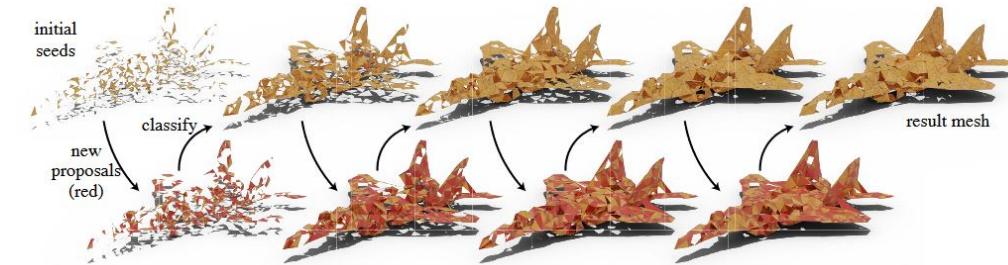


Fig. 3: Our approach iteratively constructs a triangulation by alternating between classifying triangles and proposing new neighboring triangles. Classified triangles are drawn in orange, and new proposals are drawn in red, for both only triangles with  $p > 0.5$  are shown. The input fixed vertex set is 1k points sampled from the original shape, not drawn here for visual clarity.

## [CVPR 2020] Meshlet Priors for 3D Mesh Reconstruction

スパースまたはノイジーな点群からメッシュ生成のため、ローカル幾何形状を表現するMeshletを提案。VAEでMeshletをポーズ不变な潜在空間にエンコード、点群と近いMeshlet(補助メッシュから取出)をデコード、変形の補助メッシュを利用、Meshlet間のグローバルな整合性を強める。<https://youtu.be/gIZyJ66ktog>

### Meshlet Priors for 3D Mesh Reconstruction

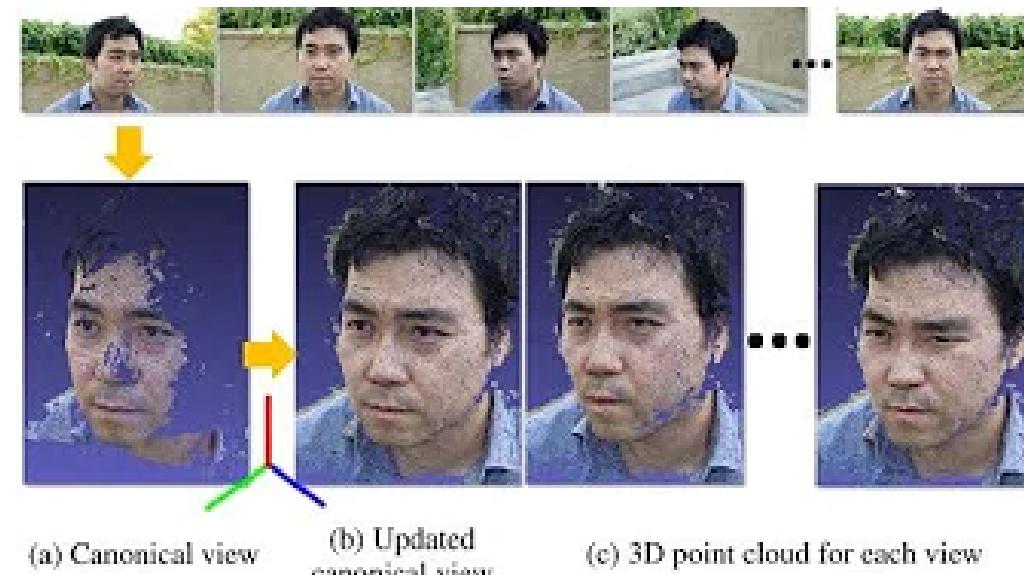
Abhishek Badki<sup>1,2</sup> Orazio Gallo<sup>1</sup> Jan Kautz<sup>1</sup> Pradeep Sen<sup>2</sup>  
<sup>1</sup>NVIDIA <sup>2</sup>University of California, Santa Barbara



UC SANTA BARBARA

## [WACV20] NRMVS: Non-Rigid Multi-View Stereo

非剛体変形下での多視点ステレオの提案. まず変形の程度の小さい画像ペアを一組決定し, 対象の基本的な3次元構造を復元. さらに, その他の変形を伴う画像群に対してもDeformation graphを利用したJoint optimizationにより, DeformationとDepthの推定を同時に行っている. <https://youtu.be/B4YBWFuYBdE>



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Visual-Inertial SLAM の Kimera。IMU の on-manifold preintegration と画像の Shi-Tomasi コーナー特徴点で VIO。DBoW2 でループ検出、GTSAM (iSAM2) でグラフ最適化。メッシュ生成と TSDF での復元。画像でセマンティッククラベリングして逆投影し、ボクセルをベイズで更新。

<https://arxiv.org/abs/1910.02490> pic.twitter.com/RSnDbyYV23

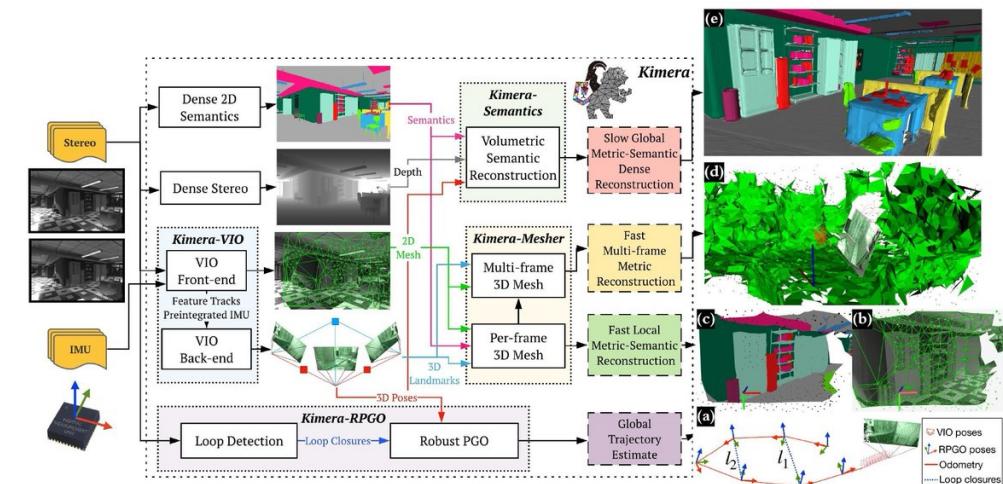


Fig. 2: Kimera's architecture. Kimera uses images and IMU data as input (shown on the left) and outputs (a) pose estimates and (b-e) multiple metric-semantic reconstructions. Kimera has 4 key modules: Kimera-VIO, Kimera-RPGO, Kimera-Mesher, Kimera-Semantics.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

複数視点の2D-CNN出力をVoxelにBack-projectして蓄積し, Voxel mapを3D-CNNに通すこと でSemantic情報を含んだMulti-view Stereoを実現. 2D-CNNにはResnet50-FPN, 3D-CNNは Skip Connectionを持つEncoder-decoderを利用. 実時間処理が可能.

<https://arxiv.org/abs/2003.10432> pic.twitter.com/e3e3J3TEua

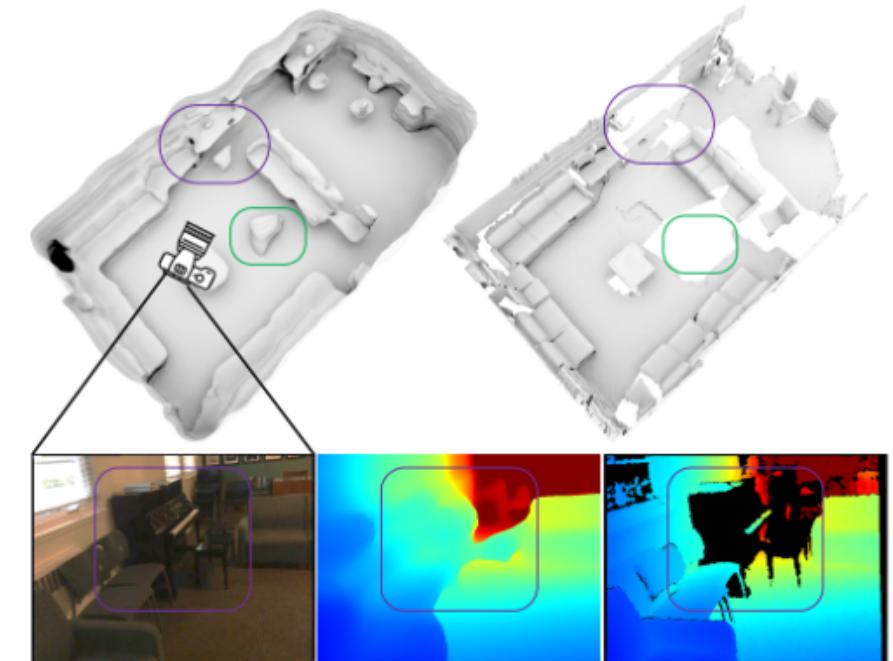
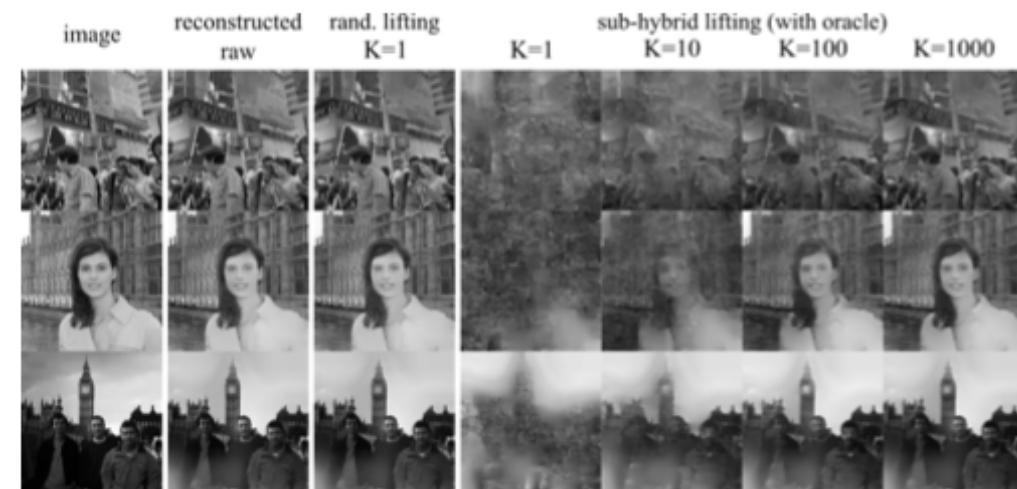


Fig. 1: Example results from our method on a Scannet scene. Top: our predicted 3D mesh and the ground truth. Notice how our method produces high quality 3D reconstructions despite not using a depth sensor. The network learns to complete geometry in a plausible way, despite it not being present in the ground truth. Bottom: a sample input image, a depth map rendered from our mesh, and the ground truth depth.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

特徴量を、それ自身を含む部分アフィン空間へ埋め込みことで、識別機能を保ちながらプライバシーアタックへの耐性を大幅に向上。部分空間同士の距離を導入し特徴マッチングを可能とした。元の特徴量と比較して、僅かな識別性能の低下により高いプライバシー保護性能を実現。

<https://arxiv.org/abs/2006.06634> pic.twitter.com/ex4qcjr200



## Li\_Through\_the\_Looking\_Glass\_Neural\_3D\_Reconstruction\_of\_Transparent\_Shapes\_CVPR\_2020\_paper.html

複数画像からガラスオブジェクトの3D形状を再構築するネットワークを提案。Visual hullで得た荒い形状を元に、各視点で屈折、反射点の法線を推論。環境マップでレンダリングした再投影誤差と、視点間を統合した点群とGTとの損失で学習。高品質な3D復元が可能などを実証。

[http://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Li\\_Through\\_the\\_Looking\\_Glass\\_Neural\\_3D\\_Reconstruction\\_of\\_Transparent\\_Shapes\\_CVPR\\_2020\\_paper.html ...](http://openaccess.thecvf.com/content_CVPR_2020/html/Li_Through_the_Looking_Glass_Neural_3D_Reconstruction_of_Transparent_Shapes_CVPR_2020_paper.html)  
<pic.twitter.com/uXXFhuXt98>

## **Wang\_VPLNet\_Deep\_Single\_View\_Normal\_Estimation\_With\_Vanishing\_Points\_and\_CVPR\_2020\_paper.html**

単一画像での法線推定手法の提案.RGB画像とマンハッタン線マップを入力とし,マンハッタン方向に沿う領域を識別するマップと法線マップをネットワークで回帰,融合する.従来手法より優れた結果を示し未見のデータに対しても推定可能なことを示した.

[http://openaccess.thecvf.com/content\\_CVPR\\_2020/html/Wang\\_VPLNet\\_Deep\\_Single\\_View\\_Normal\\_Estimation\\_With\\_Vanishing\\_Points\\_and\\_CVPR\\_2020\\_paper.html ...](http://openaccess.thecvf.com/content_CVPR_2020/html/Wang_VPLNet_Deep_Single_View_Normal_Estimation_With_Vanishing_Points_and_CVPR_2020_paper.html)  
<pic.twitter.com/vRPvFPGMEV>

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Lidarによる実世界データを用いた、従来の CADモデルによる手法よりも多彩で現実感の高い自動運転用シミュレーションを提案。Lidar 点群から動的物体や環境マップなどのアセットを作成後、物理レンダリングとDNNでドメインギャップの小さなセンサシミュレーションを行う。

<https://arxiv.org/abs/2006.09348> pic.twitter.com/nrtVaRw315

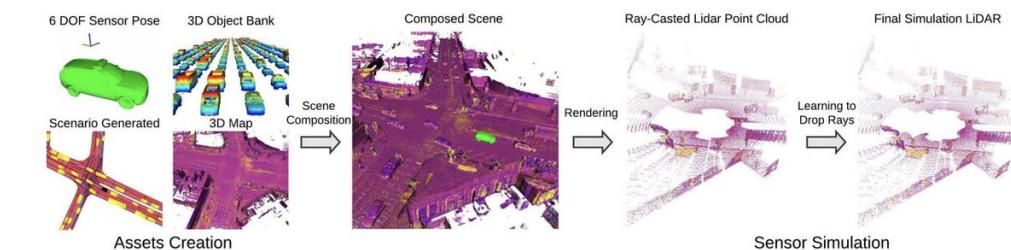
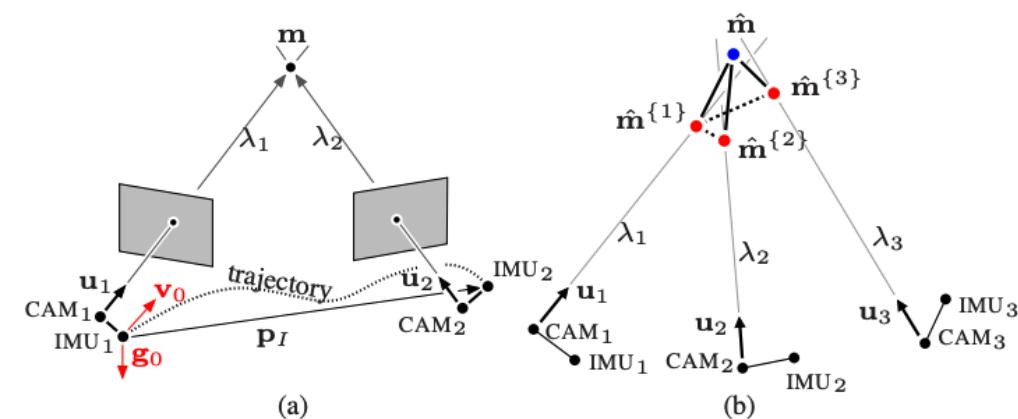


Figure 2: LiDARsim Overview Architecture: We first create the assets from real data, and then compose them into a scene and simulate the sensor with physics and machine learning.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

VIO, VI-SLAMにおける状態変数の初期化手法を提案. 効率的に不要変数を除去しつつ, 3つ以上の3D点の観測を平等に扱う新たな定式化. この線形ソルバはシンプルな構造ながら過去の手法と比較してモーション推定の精度を最大50%向上させ, 非線形ソルバの反復回数も削減.

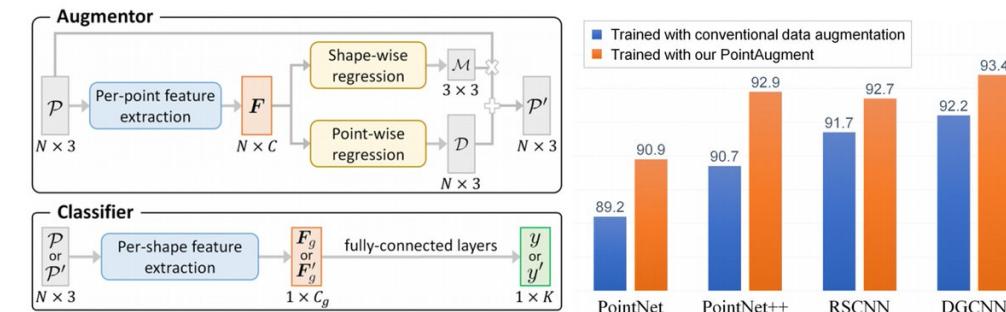
<https://arxiv.org/abs/2006.06017> pic.twitter.com/IYHoycp0k8



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

End-to-endに学習可能な点群データの Augmentorを提案. 入力点群ごとに全体の変形量と個々の点の変位量を出力し, 分類器にとってより難しい変換となるよう敵対的に学習する. 複数のモデルでランダムな水増しより良い精度を達成.

<https://arxiv.org/abs/2002.10876> pic.twitter.com/Zso2mevGNk



## S0921889019305202

LIDARスキャンを基に衛星画像での自己位置推定を行う手法を提案. 推定にはパーティクルフィルタを用い, 各パーティクル位置の衛星画像とLIDARスキャンの一致度を測るネットワークによって評価する. 衛星画像での遮蔽・陰影に頑強な位置推定が可能であることを示した.

<https://www.sciencedirect.com/science/article/pii/S0921889019305202 ...>

<pic.twitter.com/4PelxvyMge>

# Multiview LidarNet DNN: Real-Time Scene Understanding for Autonomous Vehicles.

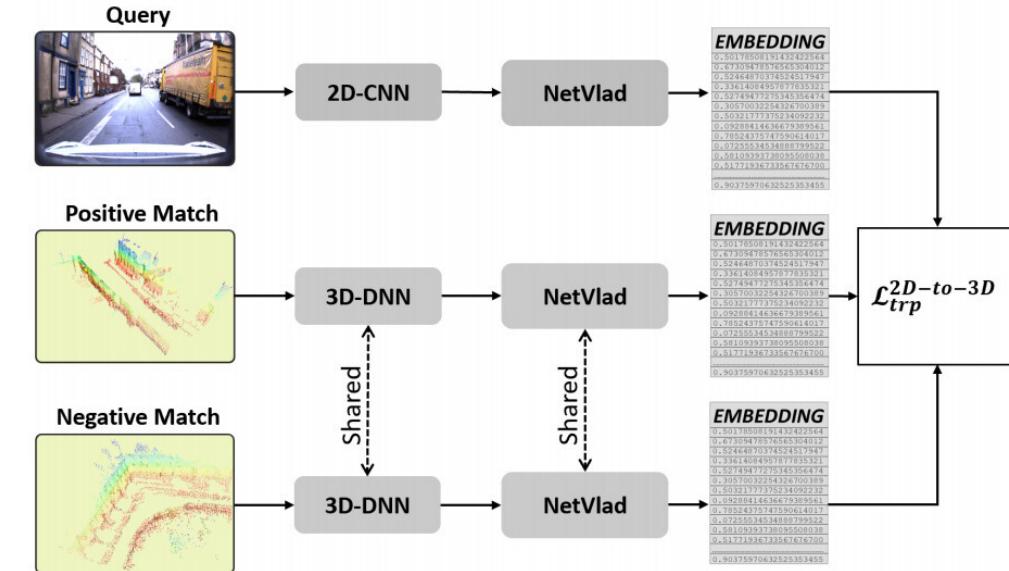
三次元点群の透視投影画像でセマンティックセグメンテーションを行い、分割結果が反応したBEV画像から物体を検出する、シンプルかつ高効率な2-stage検出手法を提案。既存手法と精度の差が大きくない上で、組み込みGPUでもマルチクラス物体検出と道路の分割を150 FPSで実現。[https://youtu.be/2ck5\\_sToayc](https://youtu.be/2ck5_sToayc)



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

広域なLiDAR地図における単眼カメラの大域位置同定. 異種のデータに対するShared embedding spaceを獲得するため, 2D-CNNと3D-DNNと一緒に学習する枠組みを提案. 同種データ内で完結するSame-Modality lossに加え, 異種データ間でCross-Modality lossを用いて学習を行った.

<https://arxiv.org/abs/1910.04871> pic.twitter.com/0Ft0zuZoH8



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

建物、部屋、物体などの関係を表すシーネグラフを動的環境に拡張し、3D Dynamic Scene Graphs を提案。Visual-Inertial SLAM の Kimera を用いてセマンティックマッピング。さらに移動物体（人のメッシュモデル）をトラッキングして、時空間の物体モデル構造を階層的に表現。

<https://arxiv.org/abs/2002.06289> pic.twitter.com/F5yOdQlh7F

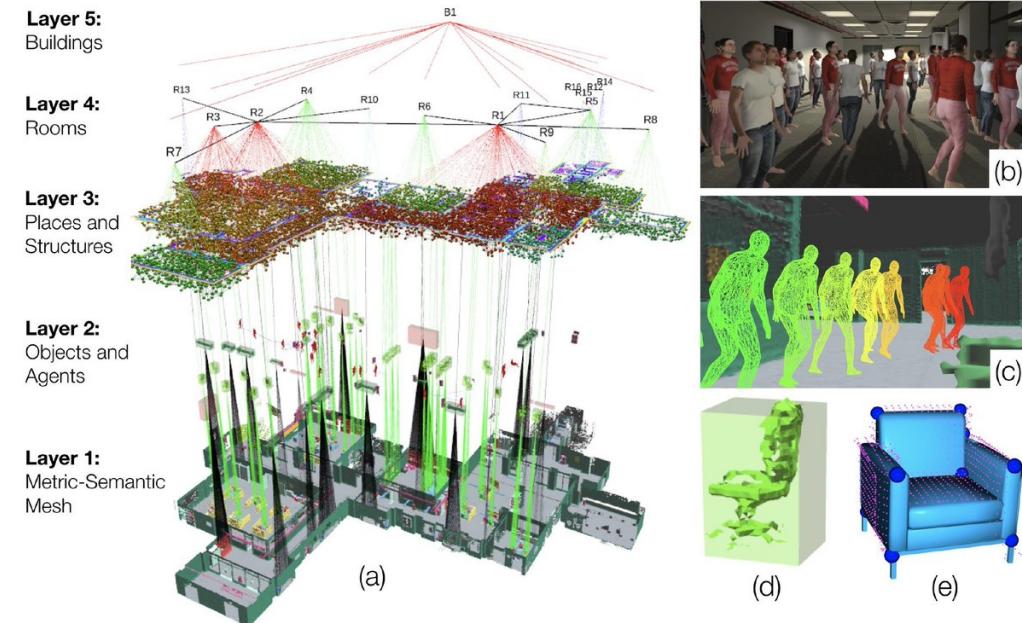


Fig. 1: We propose **3D Dynamic Scene Graphs** (DSGs) as a unified representation for actionable spatial perception. (a) A DSG is a layered and hierarchical representation that abstracts a dense 3D model (e.g., a metric-semantic mesh) into higher-level *spatial concepts* (e.g., objects, agents, places, rooms) and models their spatio-temporal relations (e.g., “agent A is in room B at time  $t'$ ”, traversability between places or rooms). We present a Spatial Perception eNgine (SPIN) that reconstructs a DSG from visual-inertial data, and (a) segments places, structures (e.g., walls), and rooms, (b) is robust to extremely crowded environments, (c) tracks dense mesh models of human agents in real time, (d) estimates centroids and bounding boxes of objects of unknown shape, (e) estimates the 3D pose of objects for which a CAD model is given.

## OccuSeg: Occupancy-aware 3D Instance Segmentation

Voxel ベースの U-Net で semantics を推定し、  
Super-voxel 間の類似度を計算して 3D instance segmentation を実現。U-Net で各 Instance に対する Voxel の占有数 (= 体積) を推定し、適切に Super-voxel をクラスタリングして Instance を生成。ScanNet Benchmark の現在 1 位。

<https://youtu.be/co7y6LQ7Kqc>



## Family 3D Photo Slideshow - 3D Photo Inpainting

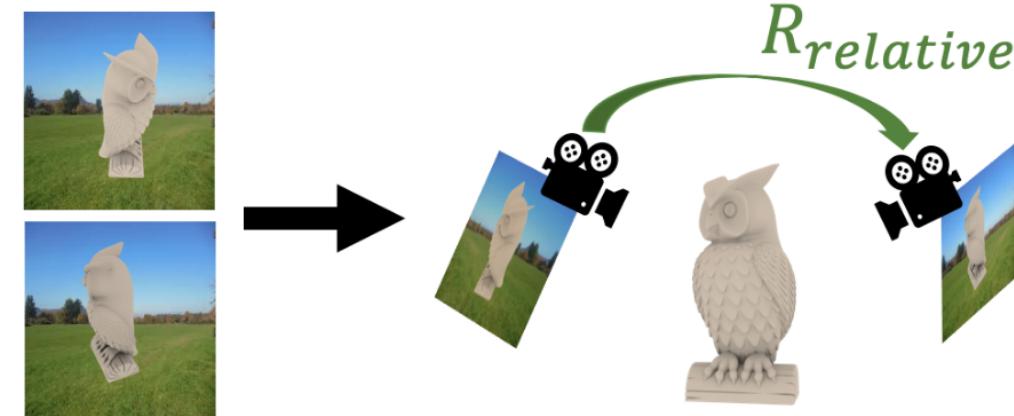
RGB-D单眼画像を入力とし，視点を変えると発生する空白領域をインペインティングするモデルの提案. Depthの断層と層状のDepth表現という着想をベースに，各層で背景を外側へ補完するようにRGB-Dを推定. Mesh表現に変換することで，エッジデバイスでも軽快に動作可能. <https://www.youtube.com/watch?v=D0JObXCfxv0> ...



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

未知物体を撮影した画像間の相対姿勢の推定。  
学習に用いられた物体以外の視点の推定は困難  
であったが 2 枚の画像を 3D 特徴グリッドにマッピングする学習を行い位置を合わせることで  
相対的な位置を推定する。学習時と大きく異なる  
物体で推論する際に従来手法より良い精度を  
示した。

<https://arxiv.org/abs/2006.03586> pic.twitter.com/SktdplZMb1



## Open4D - 01

複数カメラで撮影された動的イベントに対して、視点と時間を移動可能な4次元時空間可視化を行うシステム。シーン特化のself-supervisedなCNNを用いて静的・動的部分の抽出を行う。SfMによる既存手法で困難であった非ランバート面や、テクスチャレスな領域もキャプチャ可能に。

<https://youtu.be/sq2hhkHgtb0>

Freezing Time and Exploring Views



First Time: dense visualization despite large view changes

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

点群が平面状に分布する制約を加えたPlaner Bundle Adjustmentを提案. ヤコビアン行列のコンパクトな表現を含む新たな定式化によって精度向上と計算量の削減を両立. 評価実験で同問題設定のSOTAと比較して高速, 高精度に, そして初期値にロバストなことが示された.

<https://arxiv.org/abs/2006.00187> pic.twitter.com/rVKDwkK3LH

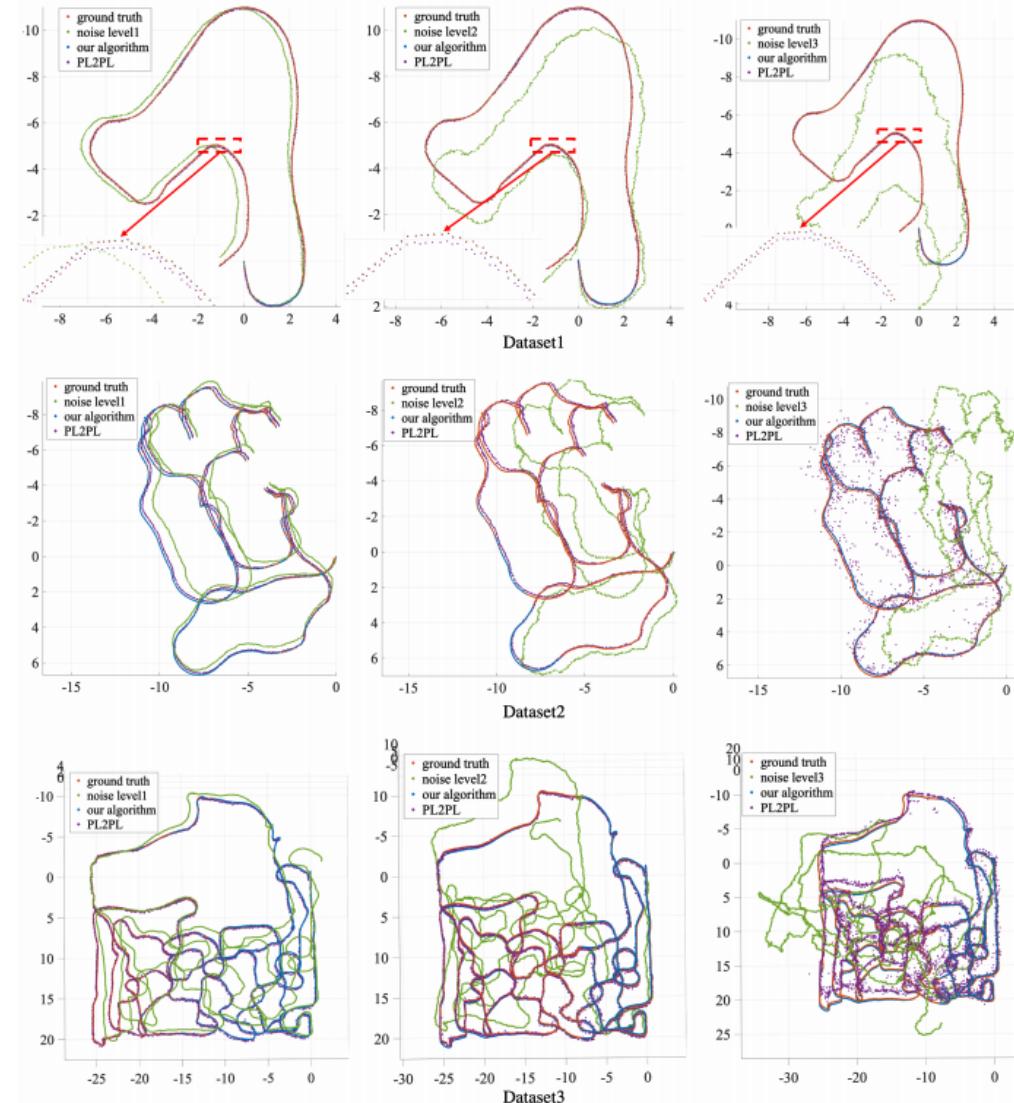
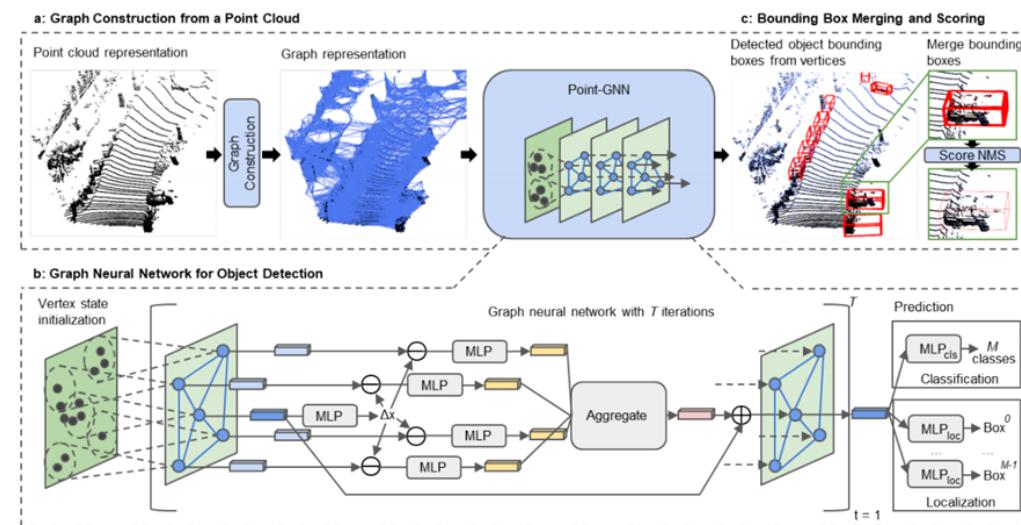


Figure 5: Experimental results of our algorithm and PL2PL [4] with varying initialization noise levels. PL2PL is not stable to large initial errors.

# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

GNNを用いた三次元点群からの物体検出手法を提案。近傍点を結んだグラフからGNNで特徴抽出し、点ごとに所属する物体クラスとBBOXを推定。最後に重複したBBOXを中心値で統合する。KITTIデータセットで従来手法を上回る精度を達成。

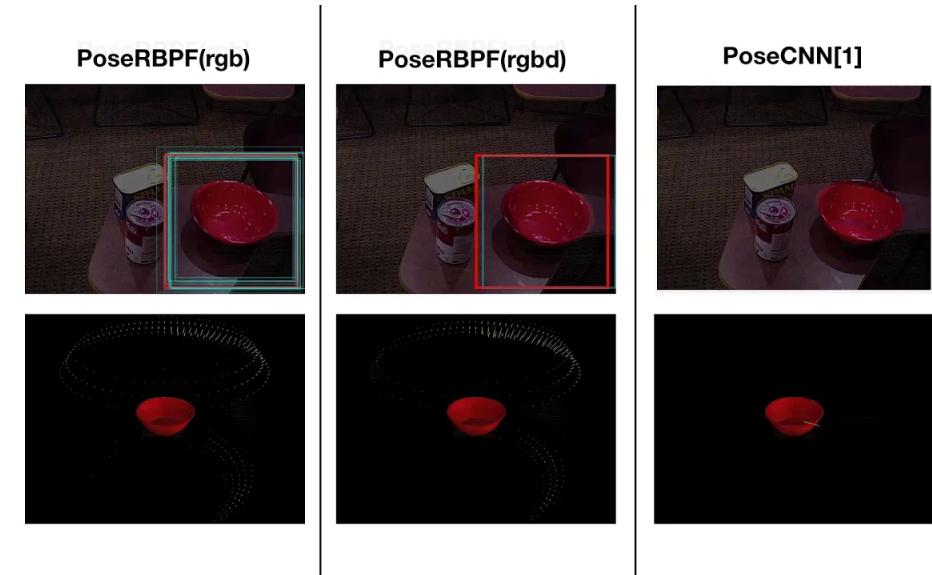
<https://arxiv.org/abs/2003.01251> pic.twitter.com/Qu1jOIW3xo



# PoseRBPF: A Rao-Blackwellized Particle Filter for 6D Object Pose Tracking

6DoF物体姿勢推定のためのRBPFを提案. 姿勢分布を分解し, 平行移動はサンプリング, 回転は物体の各回転に対するEmbeddingを予め計算しておき, パーティクルのEmbeddingをこれと比較することで評価. 6DoFを200パーティクル程度でロバストに推定しSOTA精度.

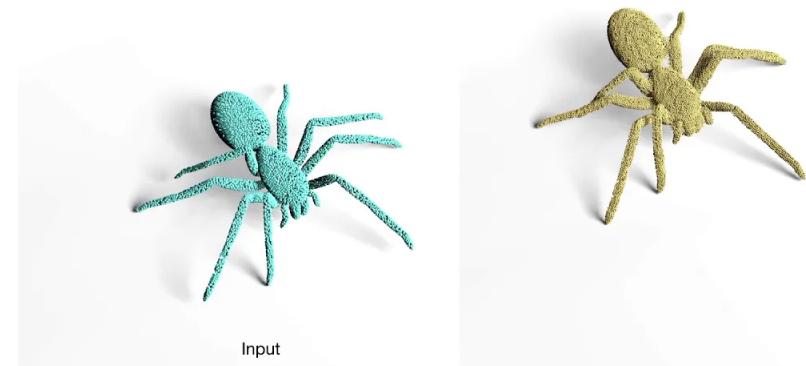
<https://youtu.be/lE5gjzRKWuA>



# Tranquil Clouds: Neural Networks for Learning Temporally Coherent Features in Point Clouds

安定的にフレキシブルな時系列点群の生成手法を提案。既存手法の安定性と多様性の両立し難い問題を改善するため、新たなTemporal Lossを導入、点群から時間的一貫性がある特徴を学習し、変形可能な数が多い点群にたしても有効性を示す。<https://youtu.be/6OoRZrqfSJ4>

Overview



## FroDO-From-Detections-to-3D-Objects.pdf

多視点のRGB画像列に基づく物体の3次元復元手法を提案. Point / Surface による相補的な形状デコードにより形状表現の効率性と記述力を両立させており, より高速な形状復元を実現している.

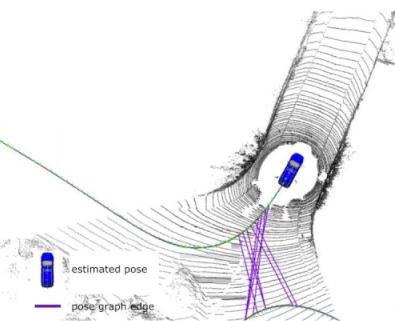
<https://research.fb.com/wp-content/uploads/2020/05/FroDO-From-Detections-to-3D-Objects.pdf> ...pic.twitter.com/Y0iFz7omo2

# RSS'2020: OverlapNet - Loop Closing for LiDAR-based SLAM by X. Chen et al.

SLAM で重要なループ検出を、位置合わせなしで end-to-end に実現。3D-LIDAR の距離画像、法線、受光強度、セマンティクスを入力。2つのスキャンの重複率とヨー角を推定。ループ拘束は SLAM 側で求める。重なりが小さくても適切にループ検出し、SuMa より高精度な地図を構築。<https://youtu.be/YTfliBco6aw>

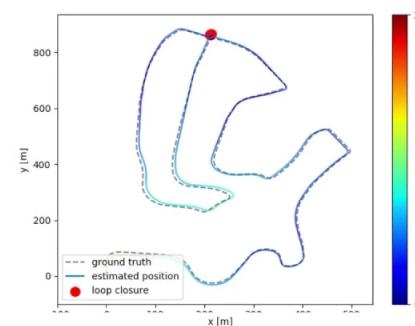
Closing Loops for 3D LiDAR SLAM

current local map



x 25

pose graph



where the car comes back to an intersection from the opposite direction.

# "A Modular Optimization Framework for Localization and Mapping" (RSS 2019 paper)

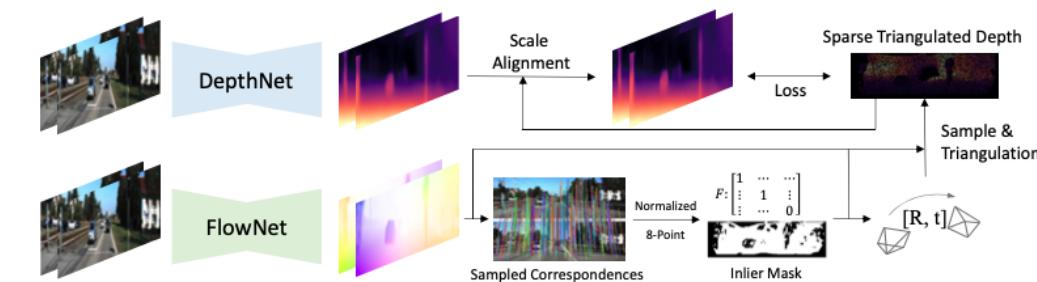
SLAMを構成要素(入出力, フロント/バックエンド, マップストレージ)に分割し, センサ種類・個数, マッピング方式(global map vs local submaps), 状態空間(SE2/SE3/SE3+vel)などの違いを包括的に扱えるミドルウェア寄りのライブラリを提案. <http://youtu.be/Bb92aMBJR44>



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

オプティカルフローを経由して8点法により直接推定した相対姿勢，さらにそこから計算した3次元点を自己教師とすることで，スケールの推定をネットワークから分離し，高い汎化性能とスケールの一貫性を実現. 屋内外のデータセットでORB-SLAMや学習ベースの手法を凌駕.

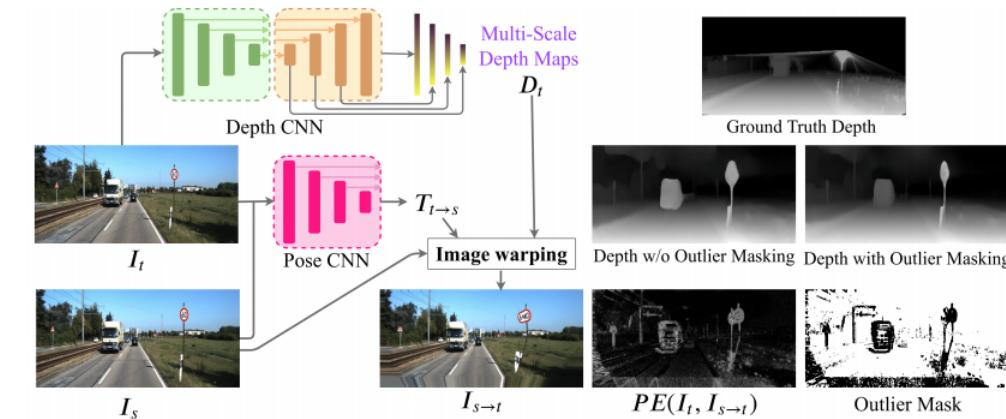
<https://arxiv.org/abs/2004.01314> pic.twitter.com/XU3nSsMtX4



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Unsupervised单眼Depth推定で精度を改善する2つの機構の提案. 特定のDepth誤りを, フォトメトリックエラーをもとにした外れ値Maskを導入し対処. また, 重み付きマルチスケール機構でアーティファクトを除去. 簡単に追加できる機構で, 他手法よりも高い精度を達成.

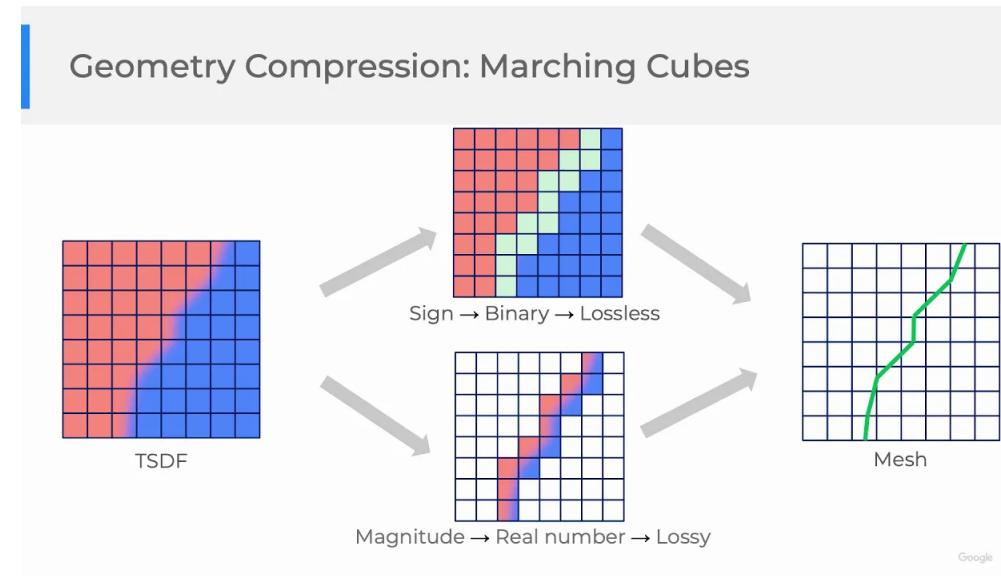
<https://arxiv.org/abs/2003.01360> pic.twitter.com/I0JA9KV4Lo



## Deep Implicit Volume Compression

Volumetricな表現で使用されるTSDFとそれに対応したテクスチャを圧縮する新しい方法を提案. End-to-Endで訓練されたニューラルネットを用い, トポロジカルなエラーを防ぐためにTSDFの符号を失わずに圧縮する. 従来手法より優れた圧縮率と歪みのトレードオフを得た.

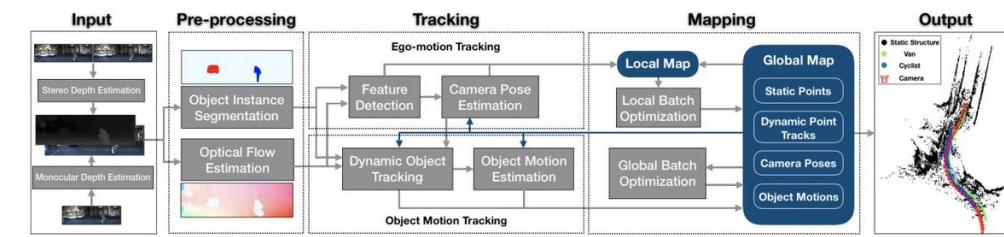
<https://youtu.be/GuLzjnFGDKs>



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

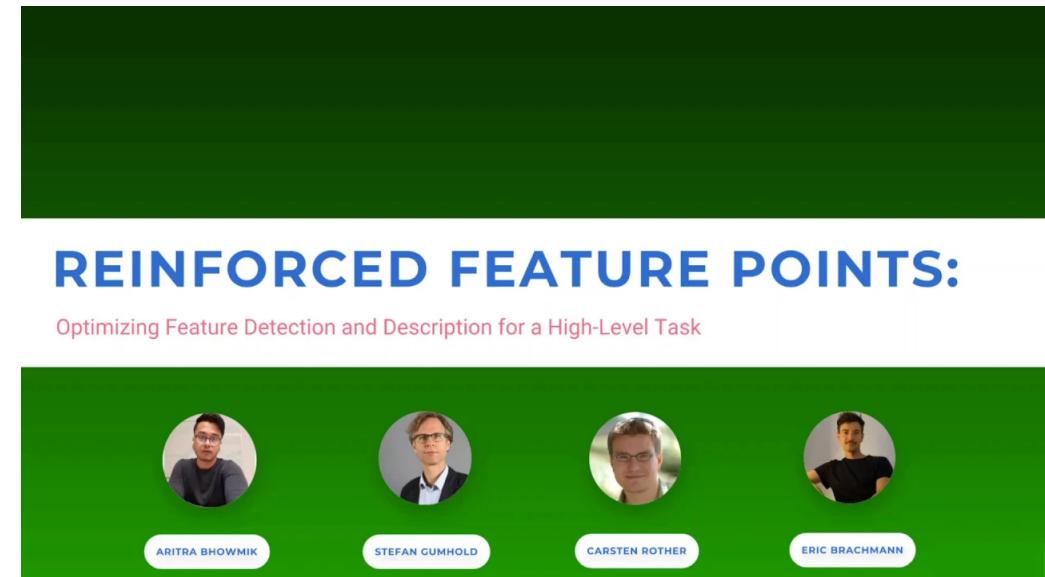
モーションセグメンテーション，動的物体追跡，カメラ姿勢，シーン剛体の姿勢変化や速度の計算を全て行い，実世界の屋外シナリオで実証可能な世界初の動的SLAMシステムを提案。ロバスト性の向上の為，カメラと物体の動きの推定はOptical Flowの改良と合わせて因子グラフ最適化

<https://arxiv.org/abs/2005.11052> pic.twitter.com/esR5t04wPi



## Reinforced Feature Points

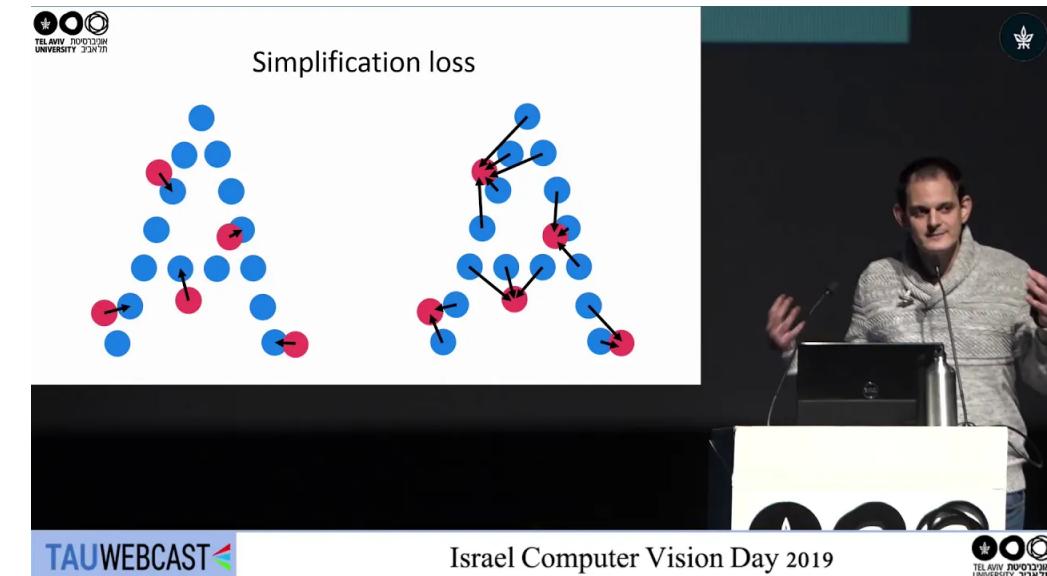
画像の特徴点検出と記述子表現をend-to-endに学習する手法を提案. 特徴点マッチングで誤差伝搬できないため, 画像間の相対姿勢誤差を負の報酬にした強化学習で特徴点検出CNNと記述子推定CNNをトレーニングする. 学習ベースの局所特徴量抽出器としてSOTAを達成. <https://youtu.be/Zttl3eDjNyc>



# SampleNet: Learning a Differentiable Point Cloud Sampling Network

微分可能な三次元点群のサンプリング手法を提案。NNによって入力点群を簡素化するDovratらの手法を拡張。最近傍サンプリングをk近傍の重み付き和で近似することで、簡素化した点群を基に入力点群をサンプリングするステップを微分可能にした。

[https://www.youtube.com/watch?v=JHz\\_lmeI8HE](https://www.youtube.com/watch?v=JHz_lmeI8HE) ...



## 6D Object Pose Estimation

RGB-D画像から特定の物体の6DoF姿勢を3段階で推定。1.2D物体認識で対象物を含む点群を抽出。2.PointNetを利用し詳細な物体抽出と並進量を推定。3.並進後、回転量をPointNetで推定。回転量推定の学習には、各点に附加した方向ベクトルが真値の方向になるように学習。

<https://youtu.be/a5JWe6mOAEs>



## image only

画像からgeometric、 semantic、 textureのVote特徴を抽出、三次元点群のVote特徴と融合し、3D物体検出の手法を提案。Multi-modalデータ融合を改善するにmulti-towerとgradient blendingの構造を使用し、SUNRGB-Dで既存SOTAより5.7mAPの精度を向上させ、SLAMようなSparse点群に対する有効性も確認。

[pic.twitter.com/6zfURbMEPw](https://pic.twitter.com/6zfURbMEPw)

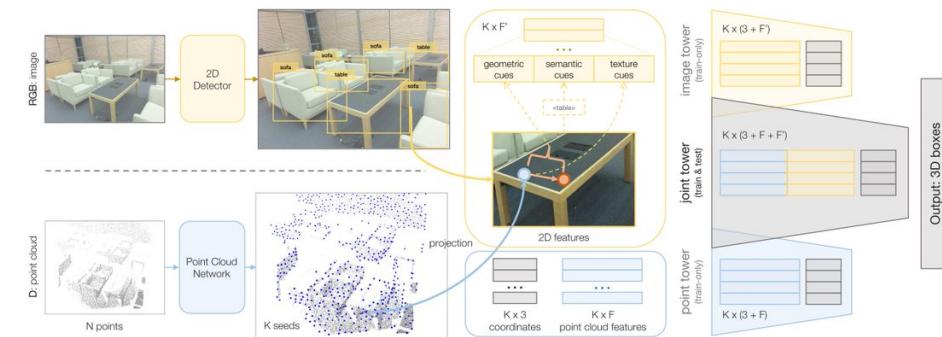


Figure 2. **3D object detection pipeline for IMVOTENET.** Given RGB-D input (with the depth image converted to a point cloud), the model initially have two separate branches: one for 2D object detection on the image and the other for point cloud feature extraction (with a PointNet++ [34] backbone) on the point clouds. Then we lift 2D image votes as well as semantic and texture cues to the 3D seed points (the fusion part). The seed points with concatenated image and point cloud features then generate votes towards 3D object centers and also propose 3D bounding boxes with its features (the joint tower). To push for more effective multi-modality fusion, we have two other towers that take image features only (image tower) and point cloud features only (point tower) for voting and box proposals.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

RGB-Dカメラ等で計測された3次元点群における幾何プリミティブ(円柱, 円錐, 球)の検出.

Point Pair Feature に各幾何プリミティブの形状に応じた拘束を導入, またLinear interpolation votingを幾何プリミティブ用に改良し, ハフ変換による低計算量での頑健な検出を実現.

<https://arxiv.org/abs/2005.07457> pic.twitter.com/AylPuLfTAw

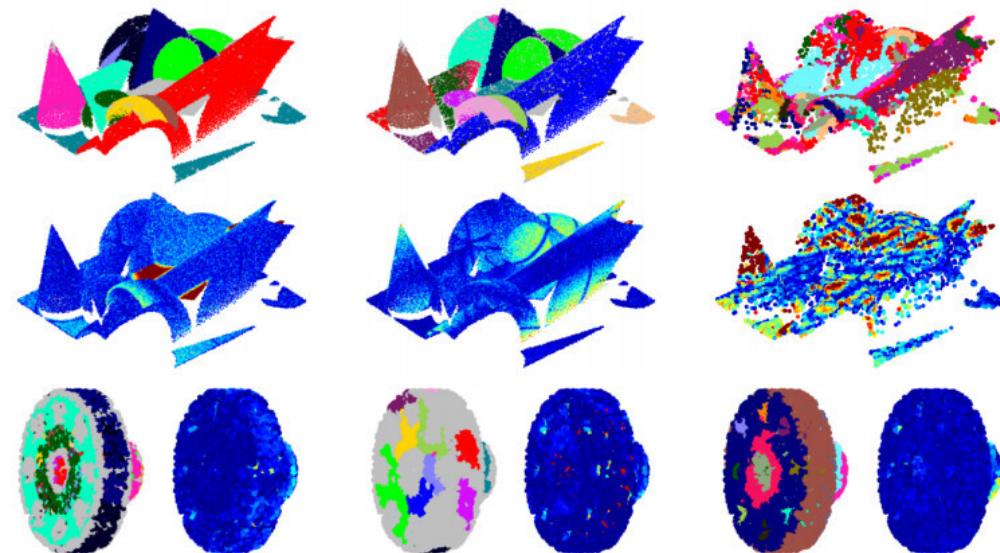


Fig. 4: (left) Our method, (middle) Efficient RANSAC and (right) SPFN results on example point clouds from the *primitect* (first two rows) and *traceparts* (third row) datasets. We show detected instances (gray: unassigned) and distances from each point to its closest primitive (blue: 0, dark red: 0.1): our method has consistently smaller distances to the detected objects than Efficient RANSAC. Being trained for object part detection, SPFN does not produce meaningful results on *primitect*. The concave cylinders on the *traceparts* data are not detected by our method.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

二つの点群について、特徴空間上で並進・回転に対するヤコビアンを数値微分で求めLucas-Kanade法で位置合わせを実行。また点群に対するEncoder-Decorderを構築し特徴をUn-supervisedまたはSemi-supervisedで学習を可能にした。

<https://arxiv.org/abs/2005.01014> pic.twitter.com/JZeCcd11hS

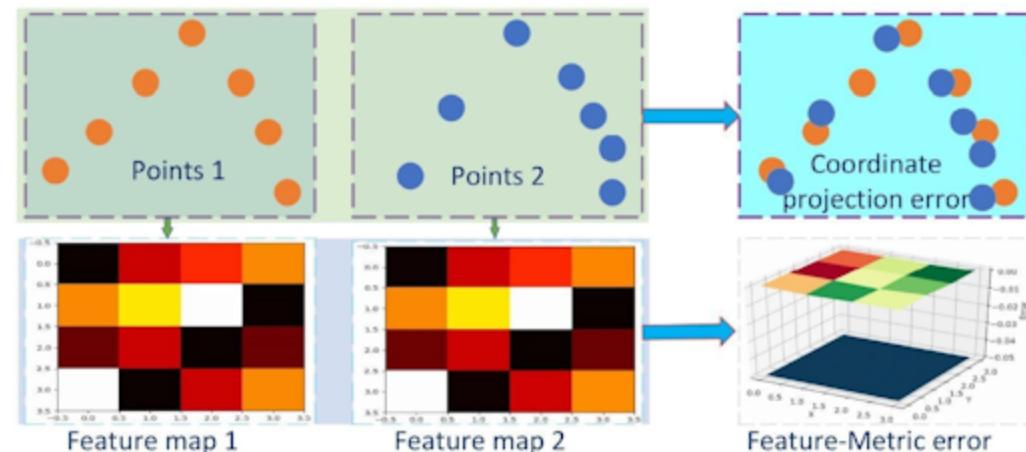
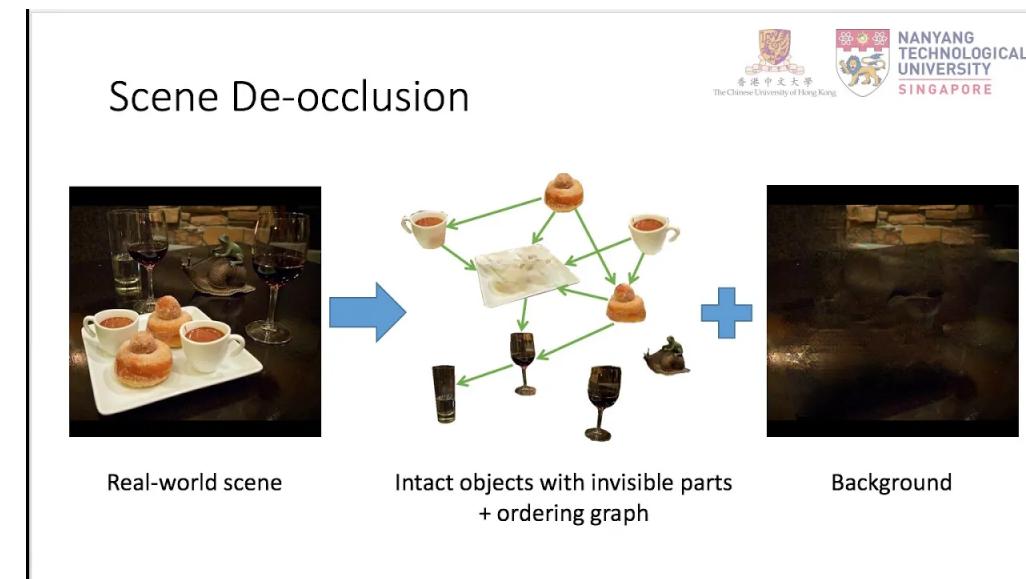


Figure 1. The difference between the coordinate-based registration (top) and feature-metric registration(bottom). The top shows the best alignment can never be achieved since this alignment has no point-point correspondences. The bottom feature-metric registration can achieve the best alignment since it optimizes the feature difference without correspondence and the feature difference is minimum in the best alignment.

# CVPR 2020 oral paper "Self-Supervised Scene De-occlusion" (click CC for subtitles)

self-supervisedに学習可能な、画像のオクレージョン領域を復元するモデルの提案。物体ごとの被オクレージョン領域推定で、増加範囲からオクレージョン関係のグラフを構築。その後推定した領域MaskからRGBを復元。教師あり学習に匹敵するパフォーマンスを達成。

<https://www.youtube.com/watch?v=xlHCyyaB5gU> ...



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

自己教師Visual Odometry手法を提案. 置み込みLSTMを利用してことで過去の経験を活かして推定をし, 未知のシーンにオンラインで適応することが可能. さらにオープンワールドでの環境の変化に対応するために特徴量を揃える手法を提案. 既存の手法を大きく上回ることを実験で確認

<https://arxiv.org/abs/2005.06136> [pic.twitter.com/YMw3oYqEK3](https://pic.twitter.com/YMw3oYqEK3)

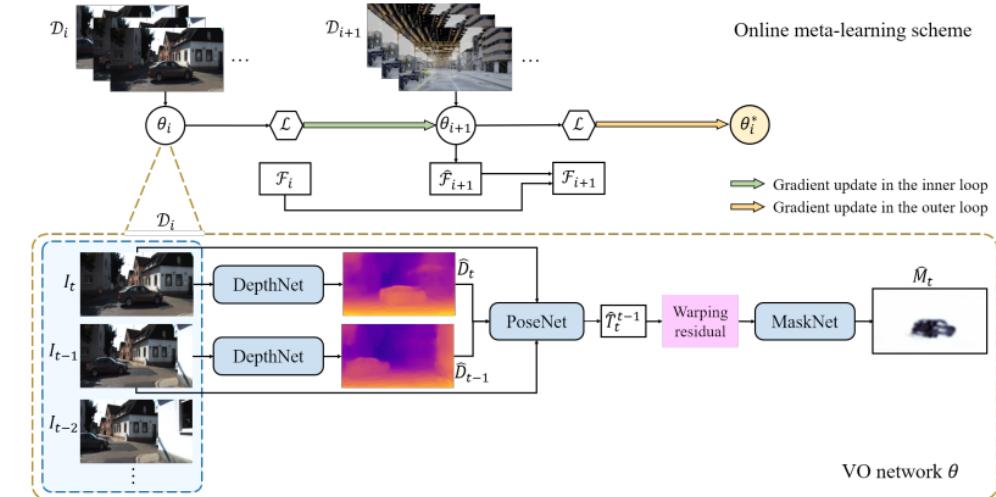
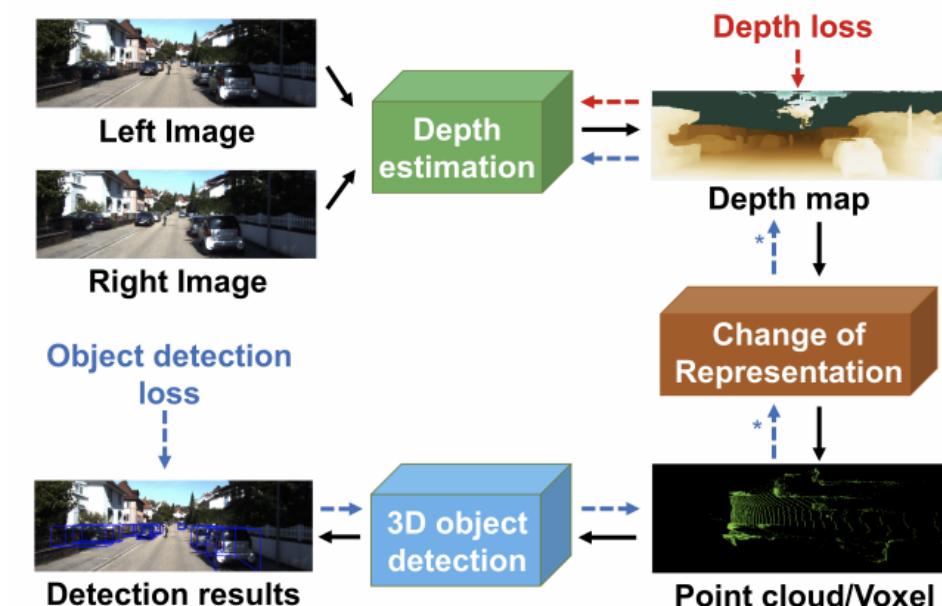


Figure 2. The framework of our method. The VO network estimates pose  $\hat{T}_t^{t-1}$ , depth  $\hat{D}_t$ ,  $\hat{D}_{t-1}$  and mask  $\hat{M}_t$  from image sequences  $\mathcal{D}_i$ . At each iteration  $i$ , the network parameters  $\theta_i$  are updated according to the loss  $\mathcal{L}$  and performs inference for  $\mathcal{D}_{i+1}$  at next time. The network learns to find a set of weights  $\theta_i^*$  that perform well both for  $\mathcal{D}_i$  and  $\mathcal{D}_{i+1}$ . During online learning, spatial-temporal information is aggregated by convLSTM and feature alignment is adopted to align feature distributions  $\hat{\mathcal{F}}_i, \hat{\mathcal{F}}_{i+1}$  at different time for fast adaptation

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

画像ベースの深度マップを擬似LiDARに変換し  
3次元物体検出するというパイプライン全体  
を, End-to-Endで学習するフレームワーク. 既  
存手法では深度推定と物体検出で別学習してい  
たが, 間の表現変化をプーリングと量子化の工  
夫で微分可能にし実現. PointRCNNと組み合わ  
せてSOTA.

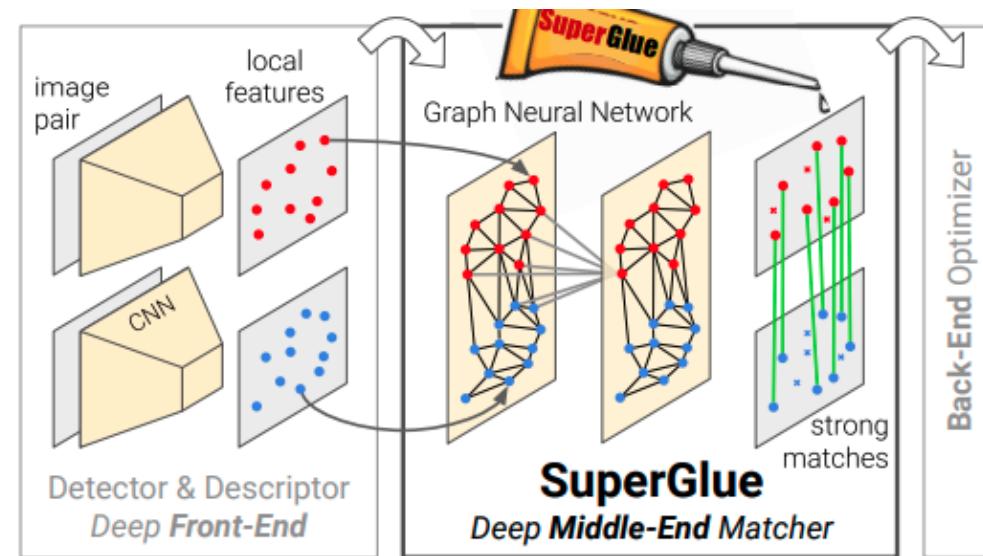
<https://arxiv.org/abs/2004.03080> pic.twitter.com/7TfZbZ079s



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

画像間の局所特徴量をマッチングするGNNを用いたアルゴリズムを提案. 視点の大幅な違いにも適用可能. 2種類のアテンション機構により画像内, 画像間でユニークな特徴量を活用する. GPUでリアルタイム動作し, 既存手法と比べ屋内外のシーンで大幅に性能向上.

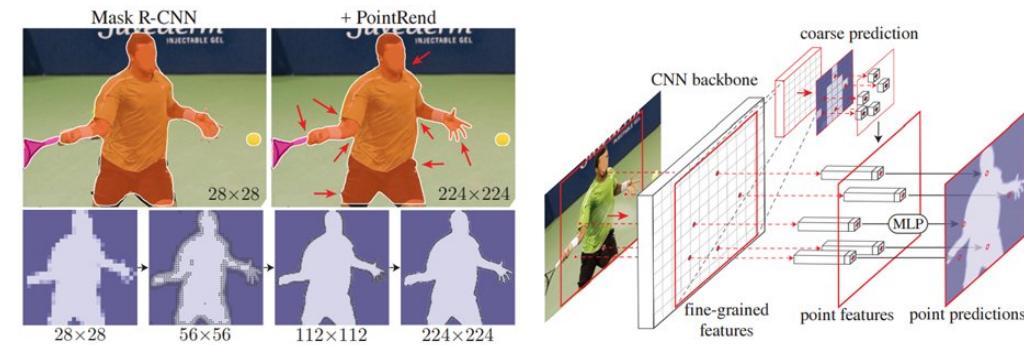
<https://arxiv.org/abs/1911.11763> pic.twitter.com/4yulTowWG7



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

CGレンダリングにヒントを得たセグメンテーションの精緻化モジュールを提案. 不確かな点をサンプリングし, MLPで推定し直すことで適応的に物体境界を精緻化. Mask-RCNNやFCNに取り付けることで, 少ない計算コストで精度向上.

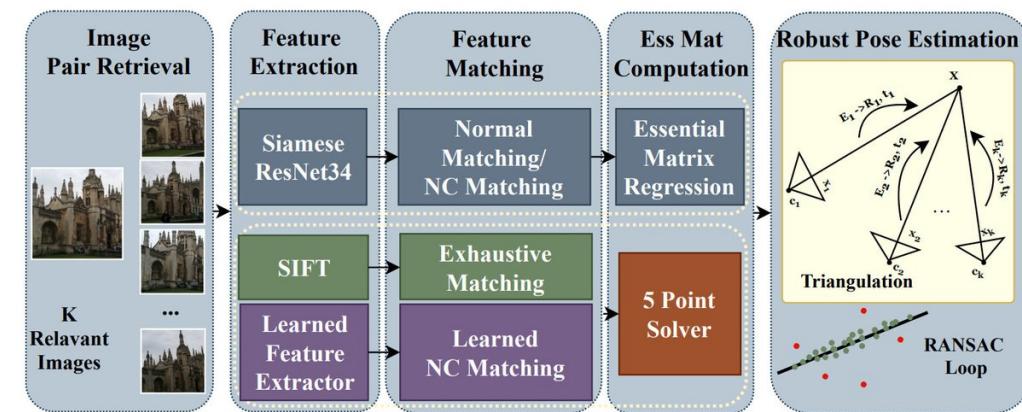
<https://arxiv.org/abs/1912.08193> pic.twitter.com/KwQ6ASEwyf



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

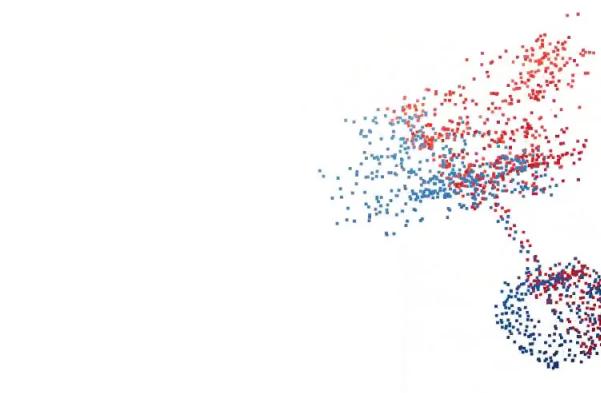
画像による自己位置推定の精度評価. Data-drivenな手法は、精度面でIndirect法等の従来手法に劣ることが通説となっている。本論文では、特徴量抽出や基礎行列計算等の各フェーズをハンドクラフトからData-drivenまで程度を変え、各組み合わせにおける精度を検証している。

<https://arxiv.org/abs/1908.01293> pic.twitter.com/Rt4IfVGCGK



## RPM-Net: Robust Point Matching using Learned Features

微分可能なSinkhornレイヤーを使い、 hybrid特徴から点と点のソフトな対応を取り、 誤対応やoverlapが少ない点群ペアでも対処できる学習ベースRobust Point Matching点群位置合わせ手法を提案。 ModelNet40での実験結果で(rule-&learning-based)既存手法より優れた性能を示す。<https://youtu.be/7hxGmMk4MZ0>



# DOOR-SLAM: Distributed, Online, and Outlier Resilient SLAM for Robotic Teams

複数ロボット上で走るStereoVOを統合する分散グラフSLAMを提案. ロボット間ループ制約の中から一貫性を保つ最大集合を探す最大クリーク問題を解いて誤検出を除去し, 積極的にループ追加を行う方針を採用. 大量のループ誤検出を除去し, 一貫した地図を生成できることを示した. <https://www.youtube.com/watch?v=h0bqURQIZGA> ...



# Voxgraph: Globally Consistent, Volumetric Mapping Using Signed Distance Function Submaps

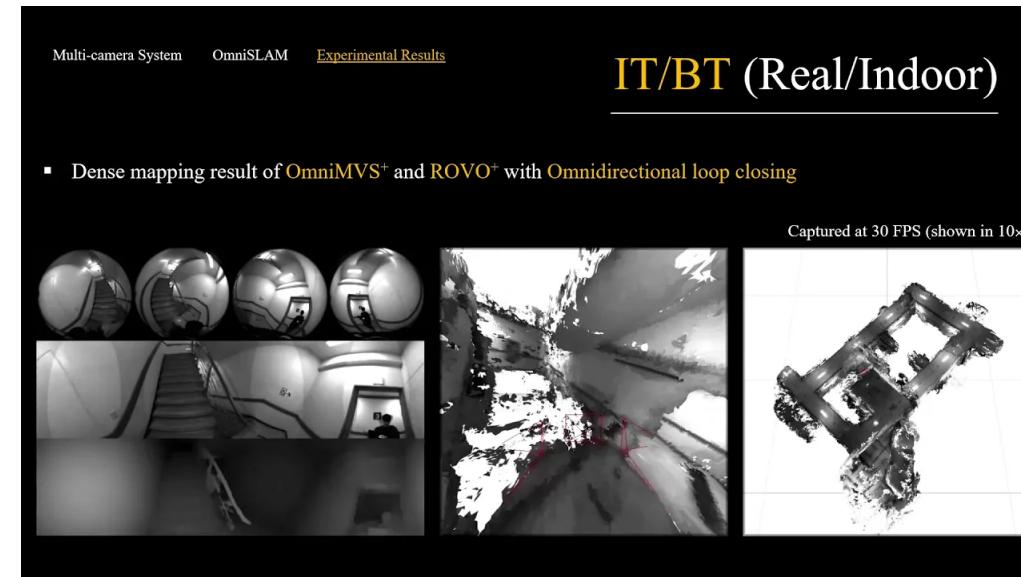
Volumetric な地図表現のグラフベース SLAM 手法。SDF サブマップの集合で環境形状を表現。SDF を利用した位置合わせで隣接拘束を生成してポーズグラフ最適化。対応付け不要なので計算コストが低い。ループ閉じ込みは、外部から DBoW などを利用してループ拘束を与える。[https://youtu.be/N9p1\\_Fkxxro](https://youtu.be/N9p1_Fkxxro)



# OmniSLAM: Omnidirectional Localization and Dense Mapping, ICRA 2020

4つの魚眼カメラを用いてロバストなオドメトリ、全方位のデプス画像生成、密な環境復元を実現。全方位のデプス画像生成にはEnd-to-Endの学習ベースによるOmniMVSを用いて生成。推定した全方位デプス画像をTSDFで統合して密な環境復元を行う。

<https://youtu.be/RFhH4j0gzsl> [pic.twitter.com/JfqoF0LKsA](https://pic.twitter.com/JfqoF0LKsA)

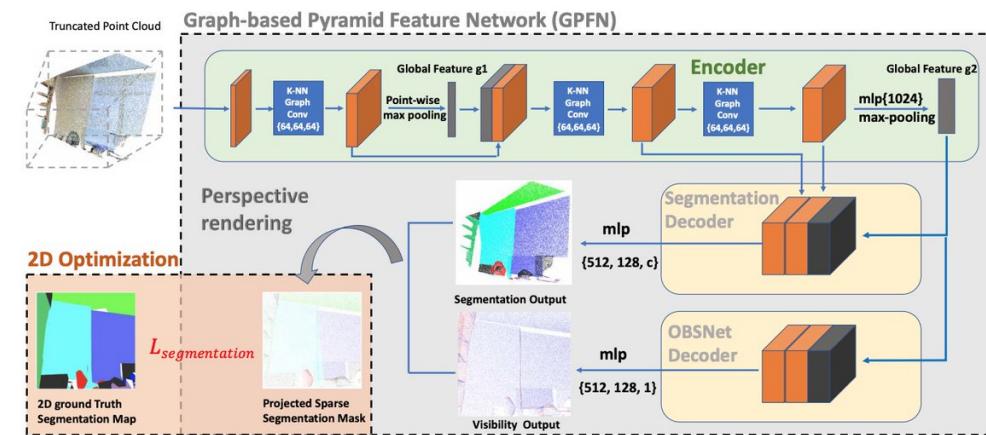


# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

3D点群のクラス推定を2Dのセグメンテーションマップのみを教師として学習するモデルの提案. 各点のクラスとvisibilityを推論し, これらで合成した2Dセグメンテーションマップと教師との損失をもとに学習する. 複数の物体を含む大規模なシーンでも高い性能を達成.

<https://arxiv.org/abs/2004.12498v1> ...

[pic.twitter.com/C4Gefp6dFl](https://pic.twitter.com/C4Gefp6dFl)



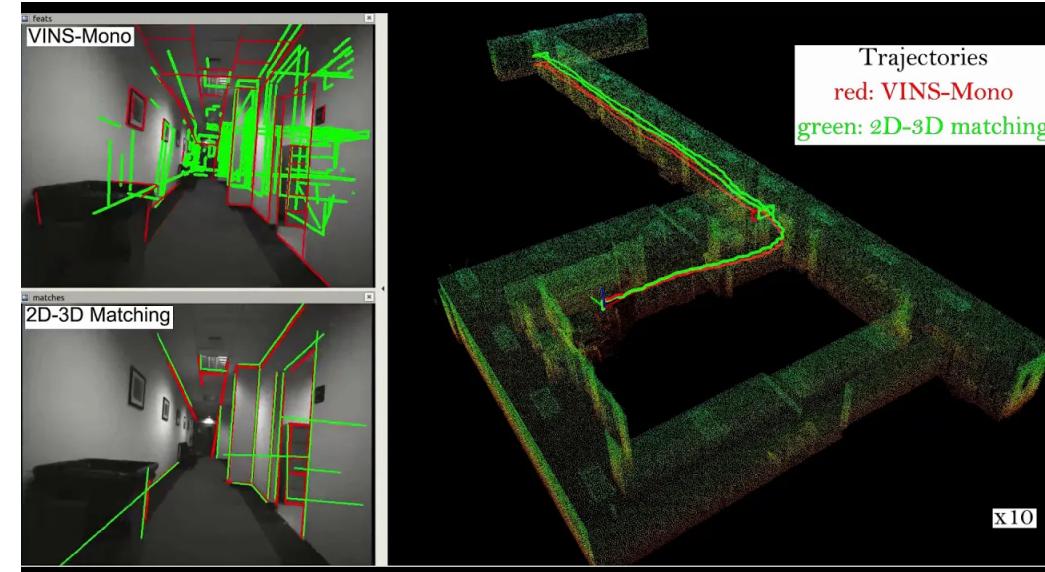
## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

单眼SLAMの初期化手法を提案.Homography推定後,Global Plane Optimization (GPO)で最適化しカメラ姿勢と平面の法線を取得.複数フレームの平面情報を組み合わせることで三角測量やHomography分解の計算負荷を減らすことが可能で,精度とリアルタイム性で優れていることを示した.

<https://arxiv.org/abs/2004.12051>

# Monocular Camera Localization in Prior LiDAR Maps with 2D-3D Line Correspondences

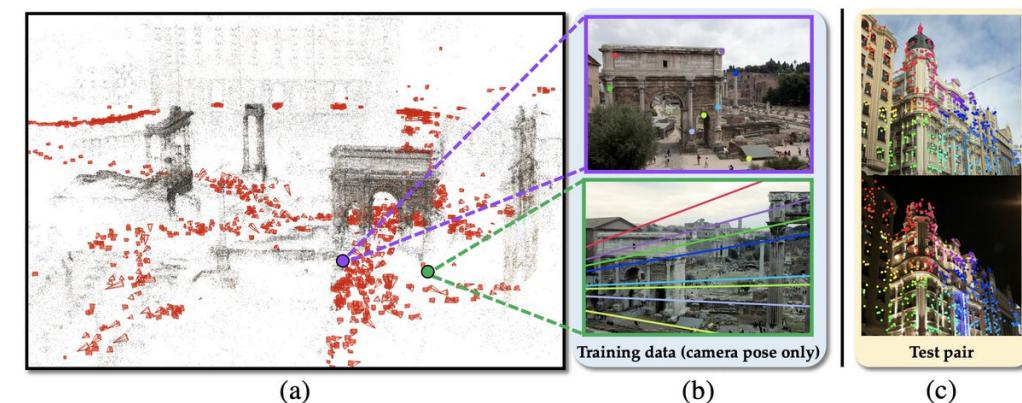
事前のLiDARマップを用いた単眼カメラ定位手法。マップからofflineで3D線を、AFMでビデオからonlineで2D線を検出。VINS-Monoからのカメラ動き予測により、2D-3D線の対応を取得。その2D-3D対応を用いたポーズ最適化により、ループクローズなしで、VIOのドリフトを低減させた。 <https://youtu.be/H80Bnxm8IPE>



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

画像間の相対姿勢のみを用いた弱教師あり学習による特徴量抽出手法を提案。中間特徴の相関に基づく微分可能なマッチング層やcourse-to-fine構造のネットワークを用い、エピポーラ幾何の拘束を組み込んだ損失関数で学習を行う。教師あり学習による既存手法の精度を上回った。

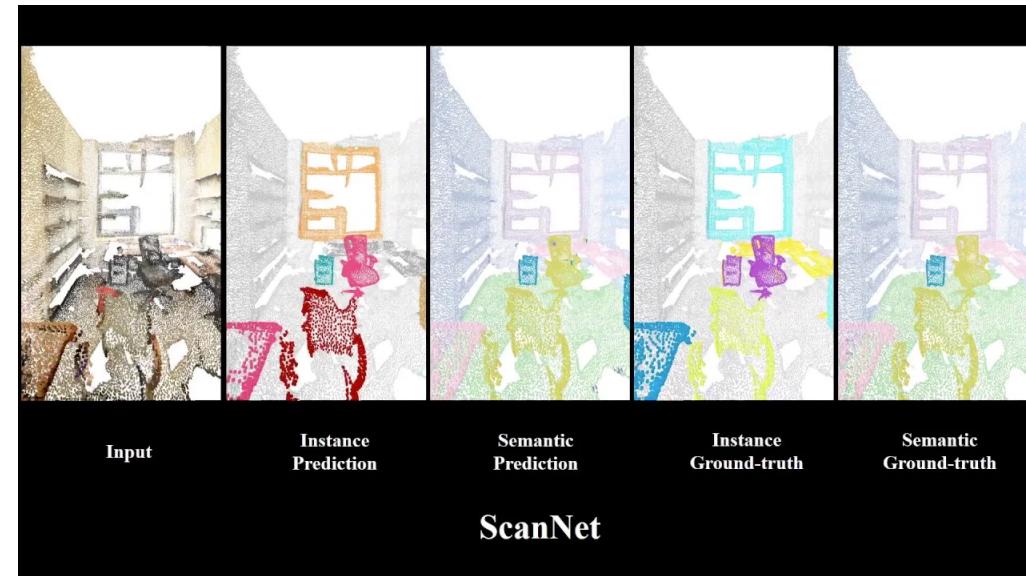
<https://arxiv.org/abs/2004.13324> pic.twitter.com/CDPfu0ZC0m



# PointGroup: Dual-Set Point Grouping for 3D Instance Segmentation

三次元点群のインスタンスセグメンテーション手法の研究。VoxelベースのU-Netで各点のクラスラベルと物体中心へのオフセットを推定。推定された座標について点群をクラスタリングすることで物体候補を生成し、後段のNNでスコアを出力する。

<https://www.youtube.com/watch?v=HMetye3gmAs> ...



## 8606275

リアルタイムでGlobally ConsistentなRGB-D SLAMのため，最適化空間を線形な特徴量部分と非線形な姿勢部分に分解. 線形部分を特徴量の二次統計量で表することで，効率的に計算を行う. 処理時間をフレーム数に対してほぼリニアな増加に抑えつつ，SOTA精度を達成.

Paper: <https://ieeexplore.ieee.org/document/8606275> ...pic.twitter.com/2VsW10Ca6T

# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

三次元点群のOne-Stage物体検出 Hybrid Voxel Network (HVNet)手法を提案。Pointwiseでの multi-scale 特徴を Hybrid Voxel Feature Extraction(HVFE)で抽出、Voxelwise attention featureにエンコード、Pseudo-Image Feature ヘデカップル、リアルタイムの31HzでSOTAを達成。

<https://arxiv.org/abs/2003.00186> pic.twitter.com/0yyB3iaOkU

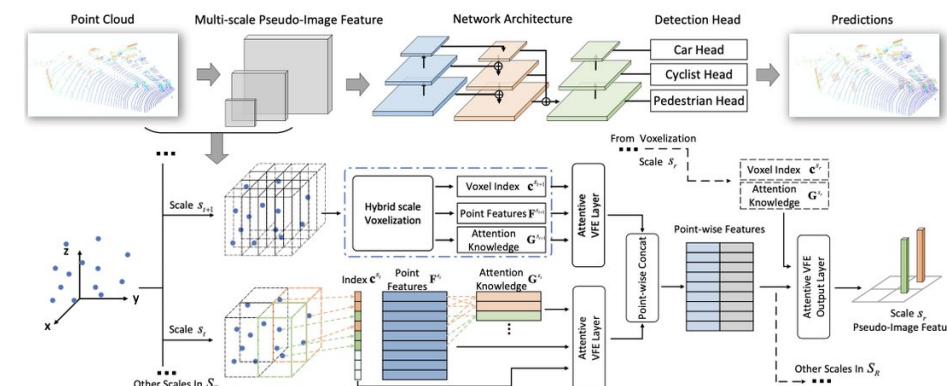


Figure 3. The first line is the whole structure of HVNet. The second line is the Hybrid Voxel Feature Extractor architecture, consisting of four parts: a) multi-scale pre-voxelization in  $x$ - $y$  plane; b) multi-scale point cloud data encoding by multi-stream AVFE layers parallelly; c) encoded features aggregation; d) dynamic feature projection by AVFEO layers with aggregated features and target scale knowledge.

## Consistent Video Depth Estimation

单眼デプス推定モデルに対して，SfMで離れた画像ペアを選択し，MVSとOptical Flowの結果から奥行と画像座標の距離を損失としてfine-tuneすることで，動画に対し一貫性のある推定を実現. 学習の前処理でMVSの結果からスケールを調整. 動物体による誤差の影響でSOTAに近い精度. <https://youtu.be/5Tia2oblJAg>



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

单眼画像からの自由空間の推定. 多視点のステレオ画像から得られた対象環境の幾何情報を单眼画像に集約し学習することで, 単一視点からでは不可視な領域に対してもTraversabilityやDepthの評価が可能に.

<https://arxiv.org/abs/2004.06376> pic.twitter.com/m7xkhPWGGu8

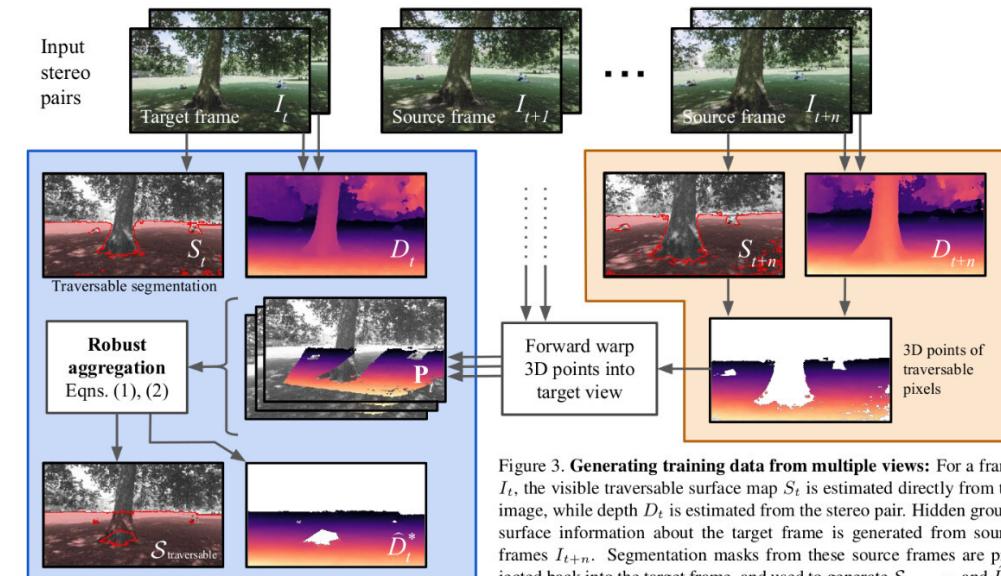
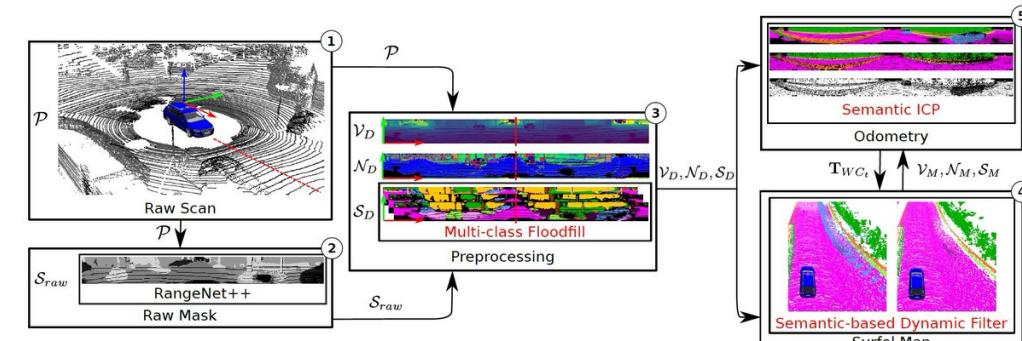


Figure 3. Generating training data from multiple views: For a frame  $I_t$ , the visible traversable surface map  $S_t$  is estimated directly from the image, while depth  $D_t$  is estimated from the stereo pair. Hidden ground surface information about the target frame is generated from source frames  $I_{t+n}$ . Segmentation masks from these source frames are projected back into the target frame, and used to generate  $S_{\text{traversable}}$  and  $\hat{D}_t^*$ .

## image only

地図を surfel で表現するグラフベース SLAM 手法の SuMa を拡張。3D LIDAR 点群を距離画像に変換し、FCN でセマンティックセグメンテーション。セマンティクスの整合性を重みとする。静止している車は位置合わせに利用される。移動している車が多い KITTI dataset の高速道路でも高精度な推定を実現。

[pic.twitter.com/M286GNv7WB](https://pic.twitter.com/M286GNv7WB)

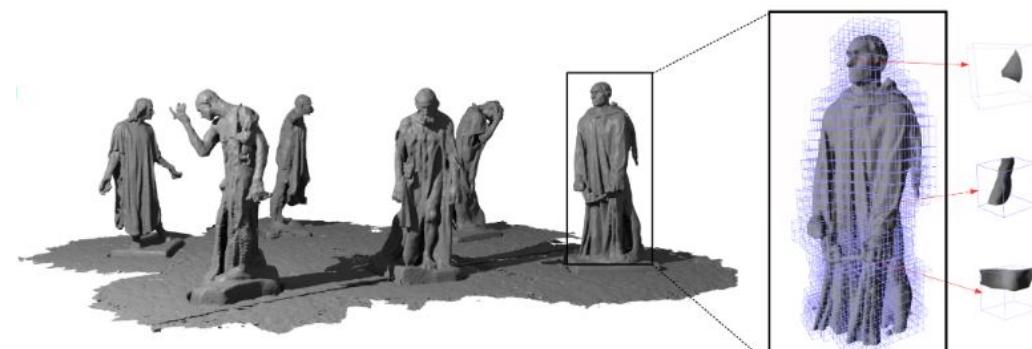


## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Kinect Fusionに利用されているTSDFを学習器に置き換えたDeepSDFを局所適用し、詳細な形状表現を可能にした。DeepSDFは全体を関数近似するのに対し、提案手法はVoxel単位で関数近似。DeepSDFが8日かかった形状復元が、提案手法では1分と大幅に短縮。

<https://arxiv.org/pdf/2003.10983.pdf> ...

[pic.twitter.com/Y89OJ9FDXh](https://pic.twitter.com/Y89OJ9FDXh)



## 1255331095656239105

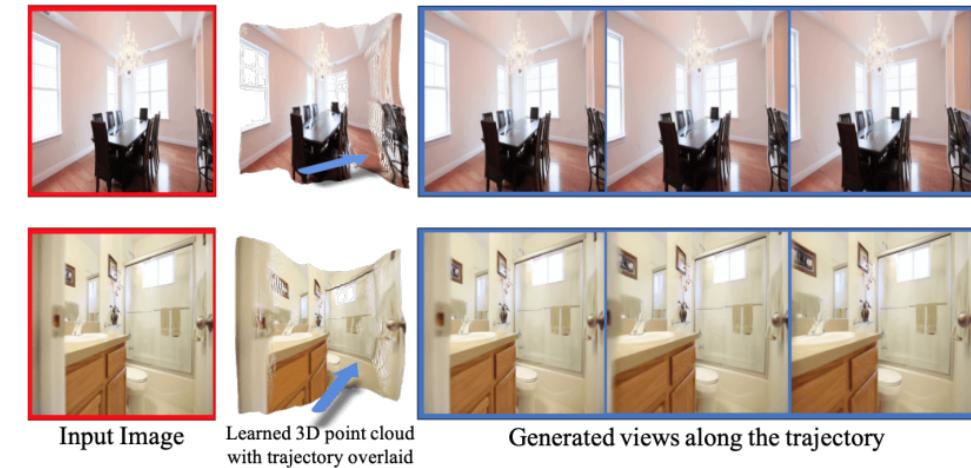
CVPR2020 Oralに採択されたStanfordのChoyさん( @ChrisChoy208 )の論文，自身が ICCV2019で提案した3D ConvNetをさらっと6Dに拡張したり，きれいに定式化して微分可能な点群マッチ提案したり，よく見たらもう1本CVPR2020 Oralに採択された自身の論文引用してるし，流石ですとしか言いようがない。

[https://twitter.com/slam\\_hub/status/1255331095656239105](https://twitter.com/slam_hub/status/1255331095656239105) ...

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

任意解像度の単一画像から任意のビューを合成するend-to-endなネットワークの提案. 推論したFeatureとDepthを用い点群を構築, 微分可能な点群レンダラーとリファインメントネットワークを通すことで欠損のないビューを合成.

<https://arxiv.org/abs/1912.08804> pic.twitter.com/cp2TcgoURr



# Online LiDAR-SLAM for Legged Robots with Robust Registration and Deep-Learned Loop Closure

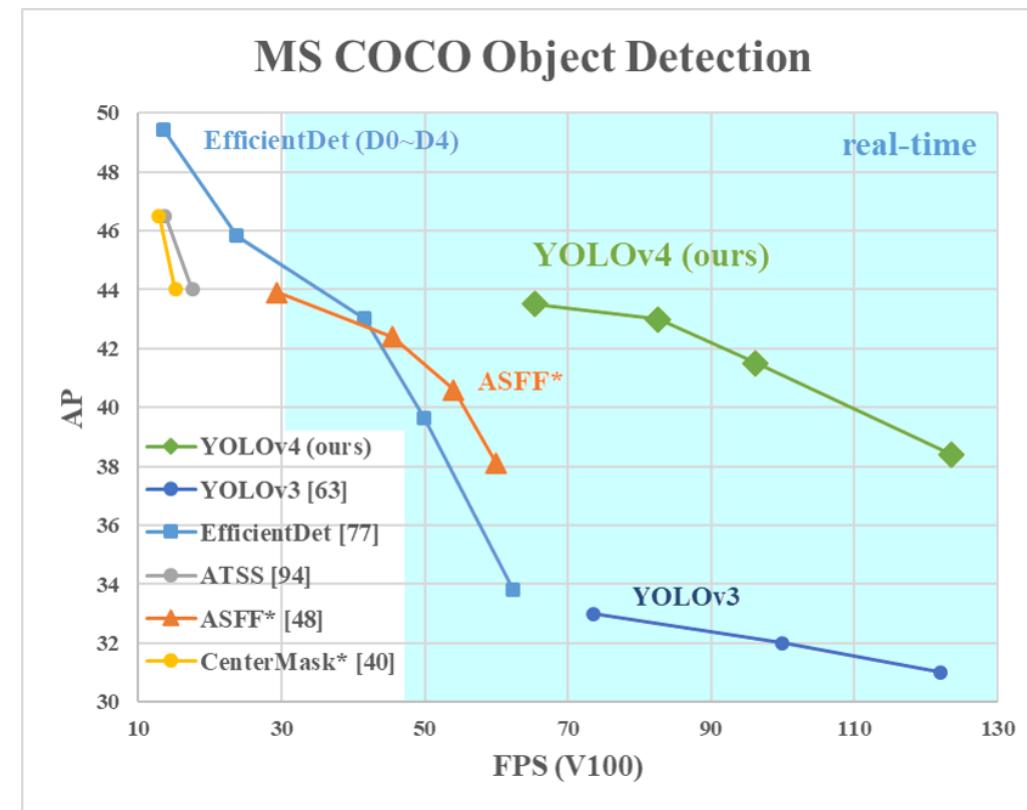
深層学習を用いた特徴量ベースのループ検出器を組み込んだグラフLiDAR-SLAMシステムを提案。四脚ロボットでも動作するように浅いネットワークを用いておりCPUで推論可能。kd-treeを用いた点群繋ぎ合わせの高速な検証方法を提案。屋内外の産業環境で口バスト性を実証。<https://youtu.be/djf7vGtf7CA>



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

CNNによる高速な物体検出器YOLOの最新版 YOLOv4を提案. 検出器の学習における, Bag of freebiesやBag of specialsによる効果を検証. バッチ正規化や残差スキップ接続など, モデルやデータセットに関して普遍的で効果のよい手法を用いることで精度を向上させた.

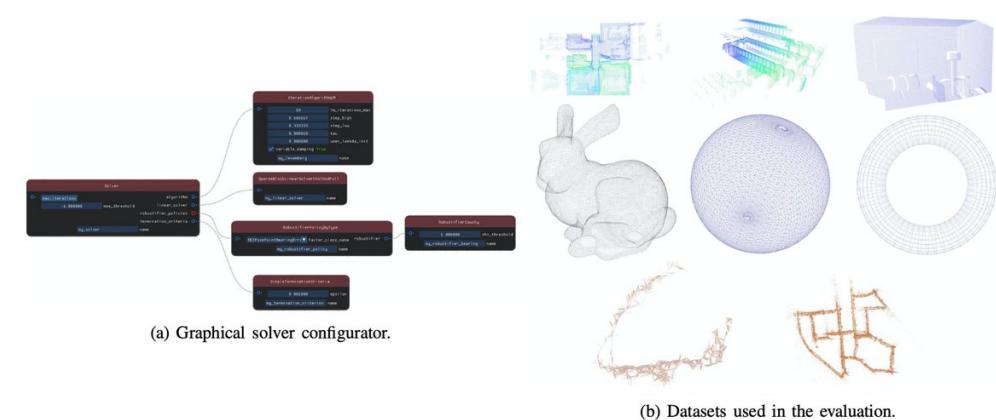
<https://arxiv.org/abs/2004.10934> pic.twitter.com/G3cdiZTGMZ



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

反復法による最小自乗を解く新たな最適化システムを提案. 既存の問題を統一的に解けるようにソルバを設計することで, 疎/密, 動的/静的な要素にシームレスに対応した. 様々な観点で比較評価を行い, 提案手法が既存システムに対し同等以上の速度, 精度性能を達成した.

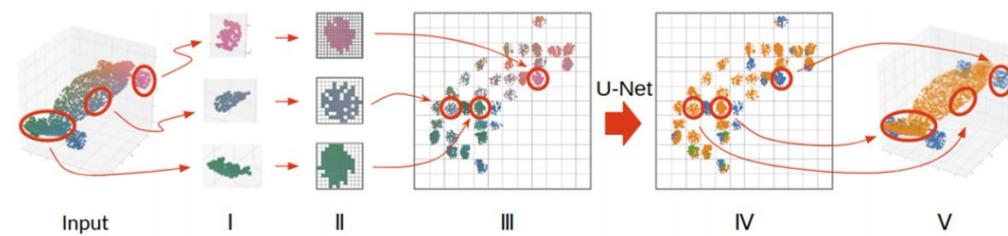
<https://arxiv.org/abs/2002.11051> pic.twitter.com/FUqWsYd55A



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

グラフ描画のアルゴリズムを用いて3D点群を2D画像に投影する手法を提案。投影した画像にU-Netを適用し、3D点群のセグメンテーションでSOTAを達成。階層的クラスタリングで得られた部分点群ごとに投影することで計算コストを削減。

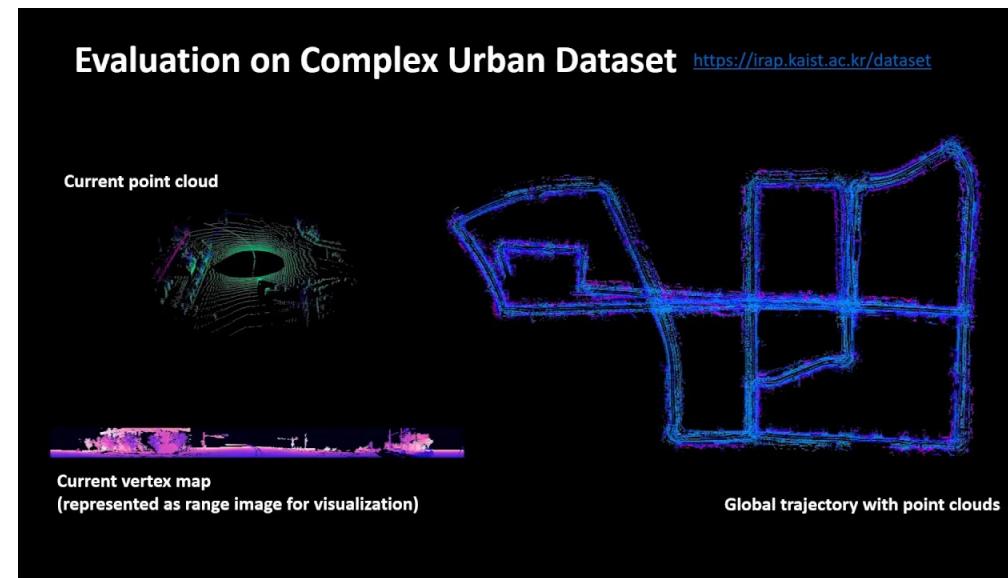
<http://arxiv.org/abs/2003.05593> pic.twitter.com/  
/4Mt0buWYu9



## Unsupervised Geometry-Aware Deep LiDAR Odometry (ICRA 2020)

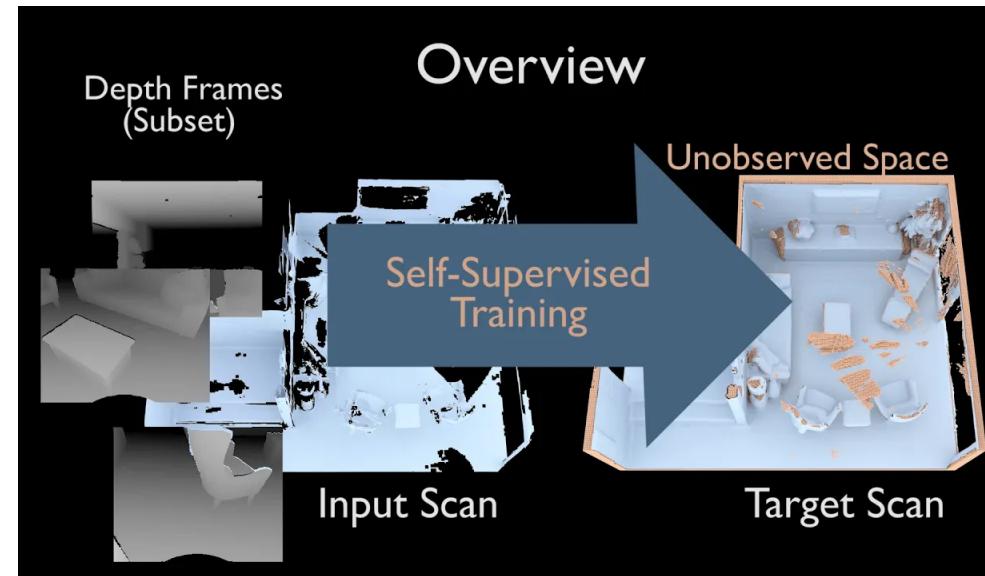
2つのLiDARスキャン間の相対姿勢推定を  
Unsupervisedに学習. 全周画像上で特徴量抽出  
や最終的なICP誤差のロスを定義することで対  
応付けを避けて学習に適した枠組みとしてい  
る.

video:<https://www.youtube.com/watch?v=-imRJXq6ZuE> ...[pic.twitter.com/WV2OHJKXzA](https://pic.twitter.com/WV2OHJKXzA)



# SG-NN: Sparse Generative Neural Networks for Self-Supervised Scene Completion of RGB-D Scans (CVPR)

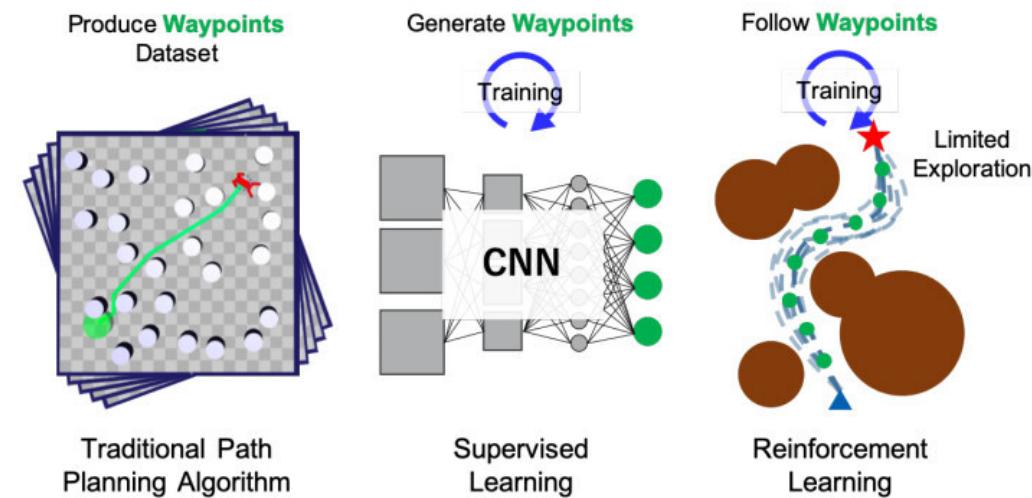
ターゲット点群の一部スキャンを削除、更に不完全な点群を入力として、自己教師ありで点群補完を学習させるSG-NNを提案。点群はスペースTSDFで表現、encoder-decoderを使って、見たこともない三次元形状にデコードでき、ターゲット点群より高い分解能点群を補完することが可能<https://youtu.be/rN6D3QmMNuU>



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

2D地図を入力とした目的地への経路・行動生成手法の提案. GOSELOと呼ばれる地図表現を介しCNNによるWaypointを生成(教師あり学習), さらにWaypointに沿うような操作量を生成する層(強化学習)を後段に追加することで安全なナビゲーションを実現している.

<https://arxiv.org/abs/2003.01641> pic.twitter.com/yskq5iMDHZ



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

RGB画像1枚から物体の3D Bounding Box(3DBB)と6DoF姿勢推定。3DBB投影時のコーナー点位置をテンプレートの重み付け和で表現し、その重みを推定する枠組み。同時に3DBBの3次元位置姿勢を推定し投影点上でのLossを定義し学習する。

<https://arxiv.org/abs/1912.07744> pic.twitter.com/IVKH8vErzB

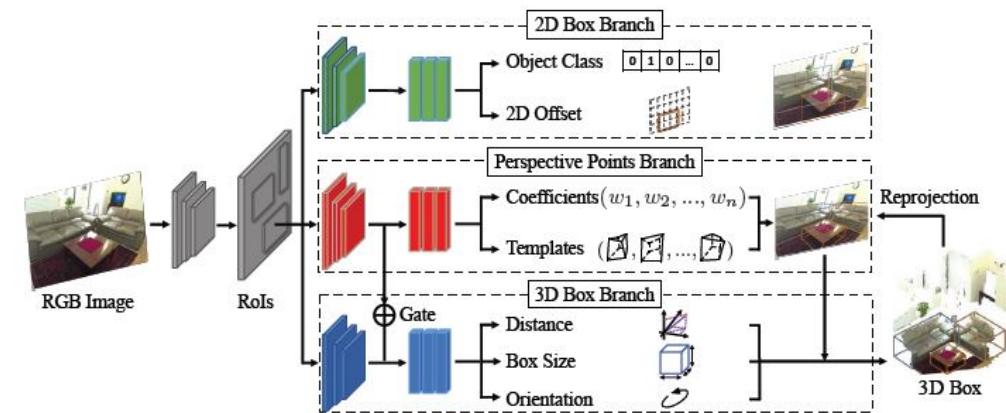
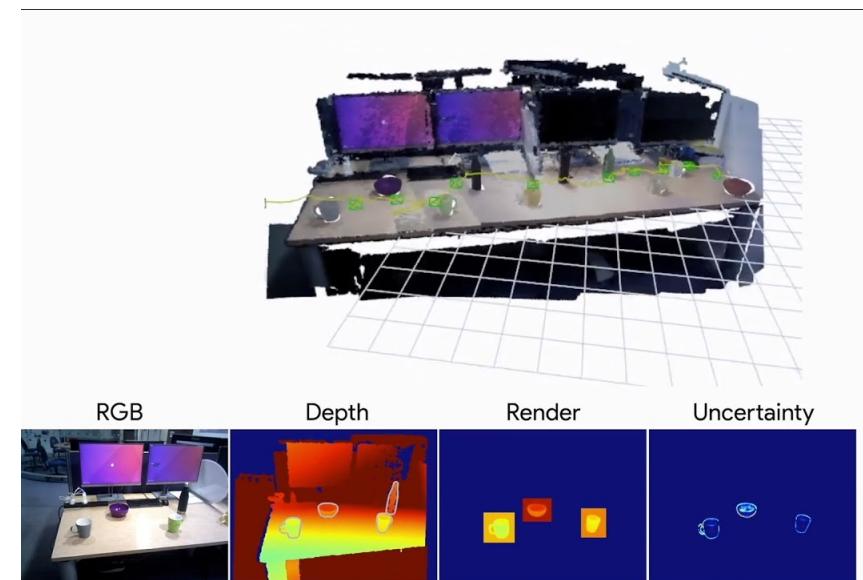


Figure 2: The proposed framework of the PerspectiveNet. Given an RGB Image, the backbone of PerspectiveNet extracts global features and propose candidate 2D bounding boxes (RoIs). For each proposed box, its RoI feature is fed into three network branches to predict: (i) the object class and the 2D box offset, (ii) 2D perspective templates (projected 3D box corners and object center) and the corresponding coefficients, and (iii) the 3D box size, orientation, and its distance from the camera. Detected 3D boxes are reconstructed by the projected object center, distance, box size, and rotation. By projecting the detected 3D boxes to 2D and comparing them with 2D perspective points, the network imposes and learns a consistency between the 2D inputs and 3D estimations.

# NodeSLAM: Neural Object Descriptors for Multi-View Shape Reconstruction

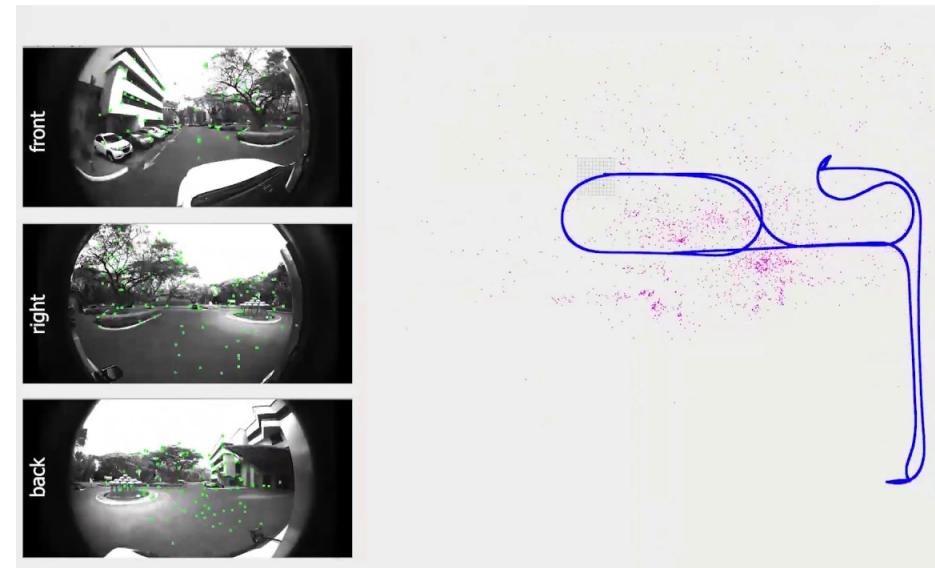
VAEを用いて学習した各物体カテゴリの3D形状特徴量を、トラッキングしながらカメラ姿勢と同時に最適化することで、単一あるいは複数視点のRGB-D画像から、オクルージョンがあっても欠損がない物体の高精度な3D形状を復元する手法を提案.

[https://www.youtube.com/watch?time\\_continue=279&v=zPzMtXU-0JE&feature=emb\\_logo ...](https://www.youtube.com/watch?time_continue=279&v=zPzMtXU-0JE&feature=emb_logo ...)



## Redesigning SLAM for Arbitrary Multi-Camera Systems

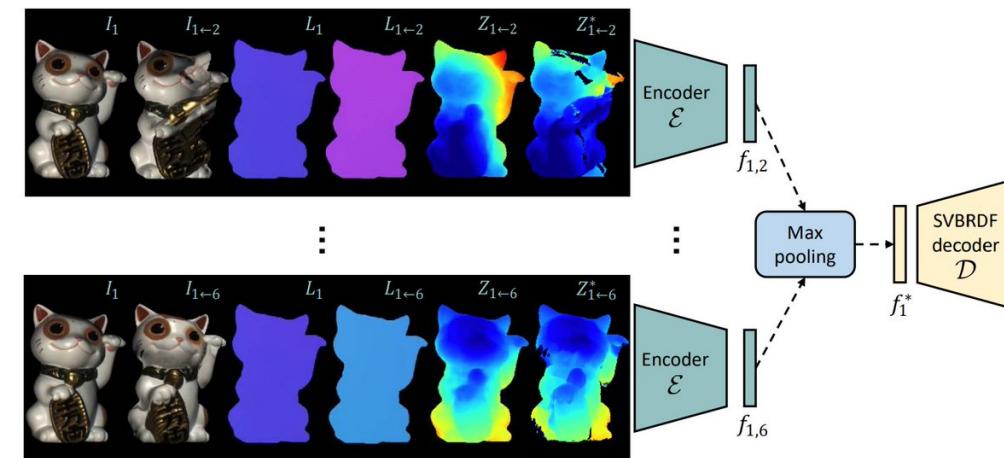
既存のVisual SLAM手法を任意の複数カメラシステムに拡張する手法を提案。適応的な初期化、センサに依存しないキーフレーム選択、voxelベースのマップ管理法を用いることで、精度を保ちセンサ固有の改良なしでの動作を実現。<https://www.youtube.com/watch?v=JGL4H93BiNw> ...



# FutureMapping 2: Gaussian Belief Propagation for Spatial AI

ワイドベースラインカメラで撮影した6枚の画像から高品質な幾何形状とSVBRDFを復元する学習ベースの手法を提案. 各画像ごとに拡散・鏡面アルベド, 法線, 鏡面粗さをネットワークで推定し, 推定結果を融合して幾何形状とSVBRDF得る. 従来難しかった疎な画像からの復元に成功.

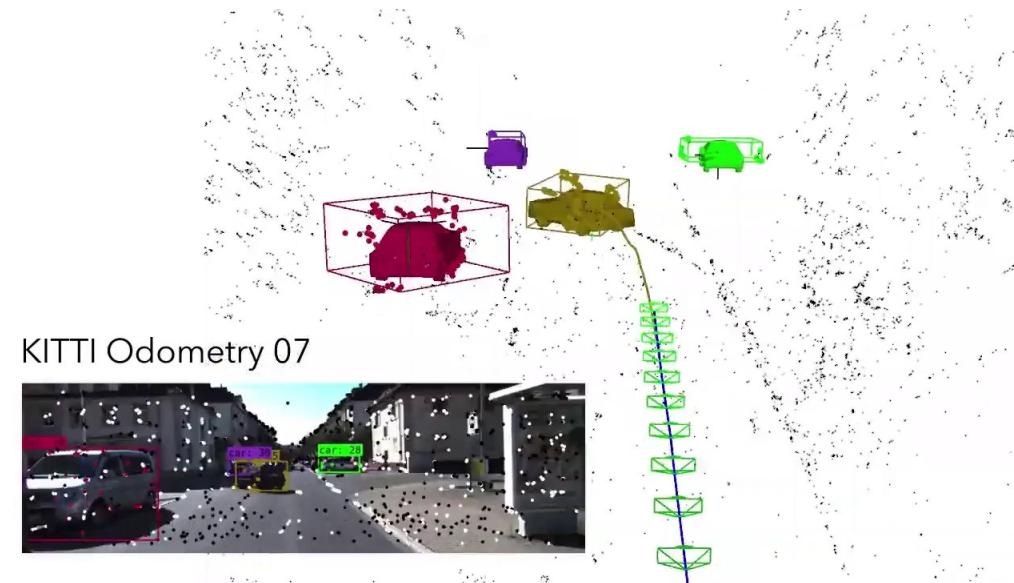
<https://arxiv.org/abs/2003.12642> pic.twitter.com/Qn5OoQkR0s



# ClusterVO: Clustering Moving Instances and Estimating Visual Odometry for Self and Surroundings

ステレオカメラを用いて，物体上の特徴点のクラスター化と，自身と物体の動きの推定を同時に行うシステムを提案. クラスター化は，物体検出による特徴点のクラスラベルを用いて，3次元位置も考慮したCRFにより実装. シーンに依存せず，オンラインでの動作を可能にした.

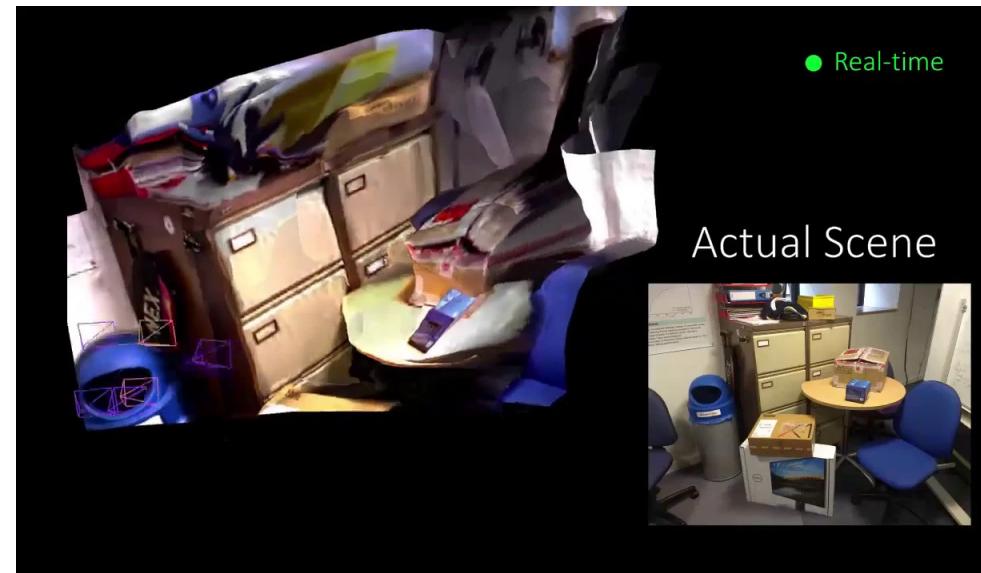
<https://www.youtube.com/watch?v=paK-WCQpX-Y&feature=youtu.be ...>



# DeepFactors: Real-Time Probabilistic Dense Monocular SLAM

CodeSLAMをベースにした新たな深層学習ベースのVisual SLAMシステムを提案。既存のコードによるコンパクトなデプスマップ表現に加え、損失関数の改善、ループクロージングと全体最適化の追加により、精度とロバストを向上。さらにリアルタイム動作を実現した。

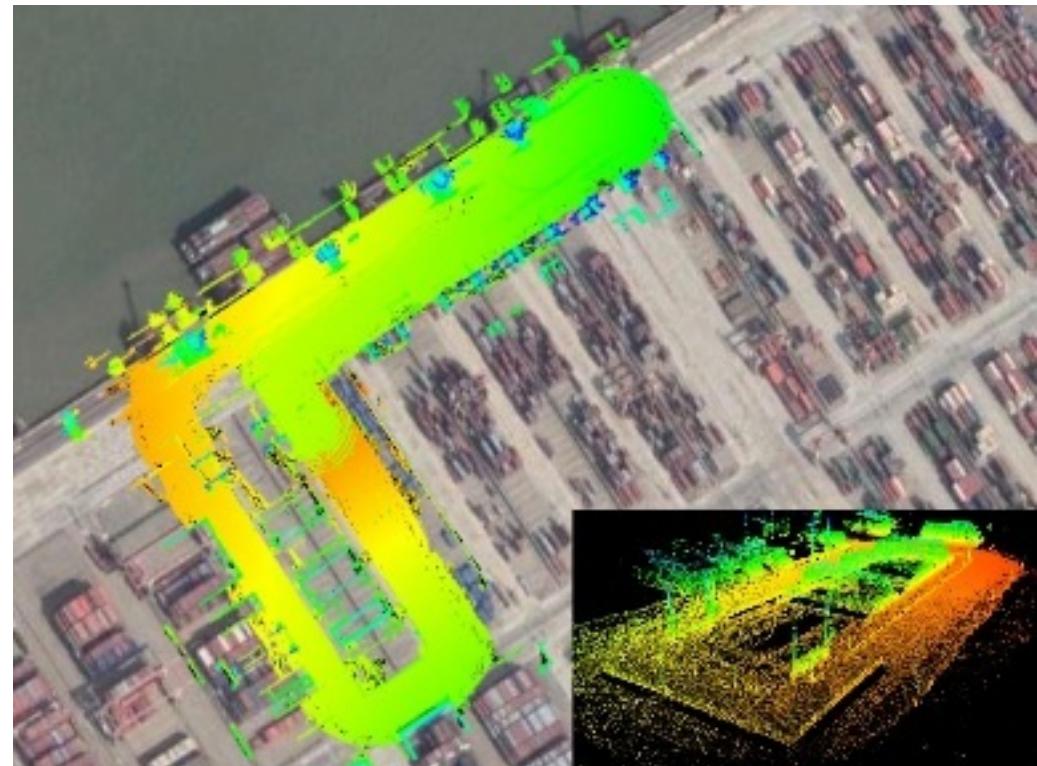
<https://www.youtube.com/watch?v=htnRuGKZmZw> ...



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

LIDAR-IMUのオドメトリ推定手法. Error State Kalman Filter上でTight-couplingに最適化を行うことで、従来のグラフベース手法と近い精度を維持しつつ大幅に処理速度を向上させた.

paper:<https://arxiv.org/abs/1907.02233> pic.twitter.com/IQ4YAzhl8q

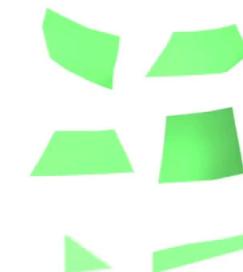


# Local Implicit Grid Representations for 3D Scenes (CVPR 2020)

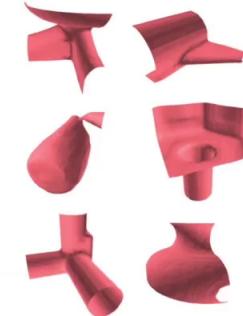
Object Levelより細かく、新たな形状表現Local Implicit Grid (LIG)を提案。AutoencoderでPartの形状をlatent vectorにエンコード、入力点群と復元ロス最小なlatent vectorを最適化、LIGでの内挿により形状にデコード、機械学習で点群からScene Levelの三次元形状復元が可能。

<https://youtu.be/XCyl1-vxfll>

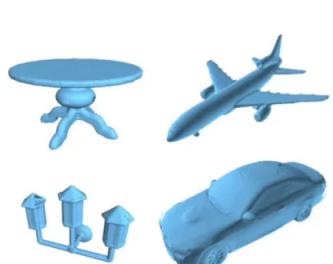
Motivation



(a) Micro Scale  
*Trivial abstraction*



(b) Part Scale  
*Generalizable abstraction*

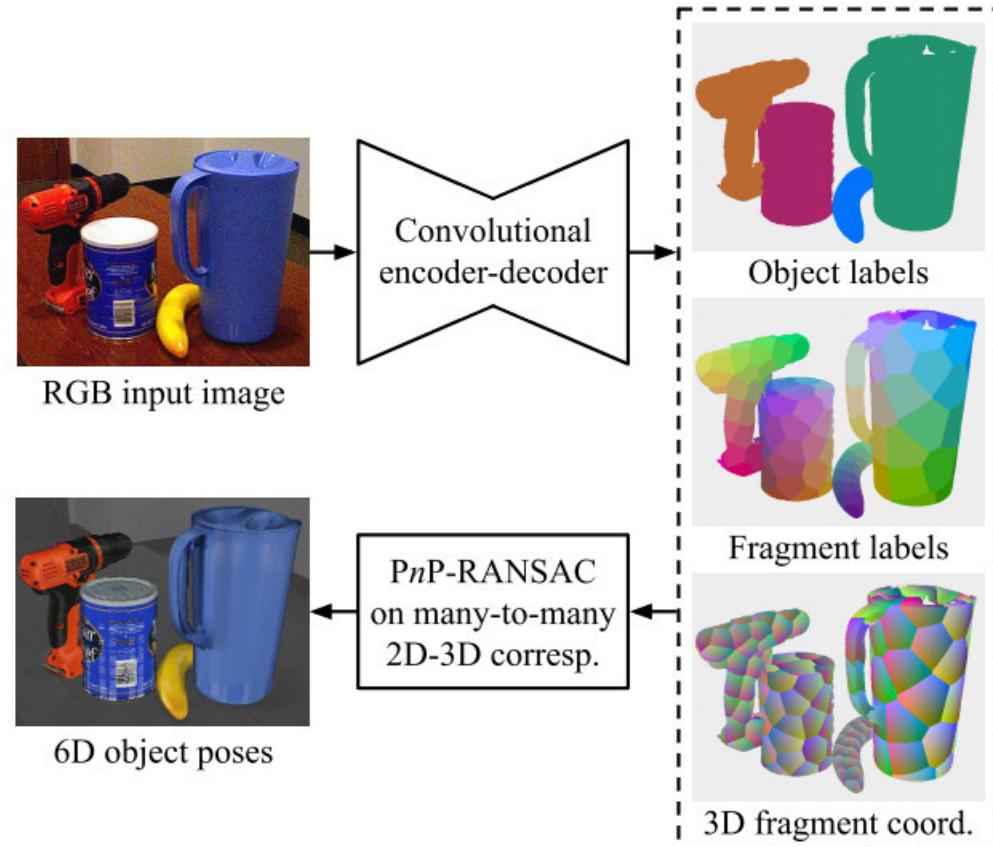


(c) Object Scale  
*Does not generalize*

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

RGB画像における物体の6DoF姿勢推定.  
Surface fragmentによる3次元モデル表現を介し、物体のPoseを各PixelがどのようなインスタンスやSurface fragmentに対応しうるかを学習。得られた多対多な2D-3D対応をPnP-RANSACによりロバスト化。

<https://arxiv.org/pdf/2004.00605.pdf> ...  
[pic.twitter.com/Qn2Etv1Nst](https://pic.twitter.com/Qn2Etv1Nst)



## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

画像の位置合わせに必要なエラーについて、輝度差に代わる新たなメトリックSGFを提案。

SGFは勾配画像のコントラストを局的に正規化し、勾配方向の内積を利用。DSOにSGFを適用したところ精度が改善。

<https://arxiv.org/pdf/2004.04090.pdf> ...

<pic.twitter.com/HL3hkOLo61>

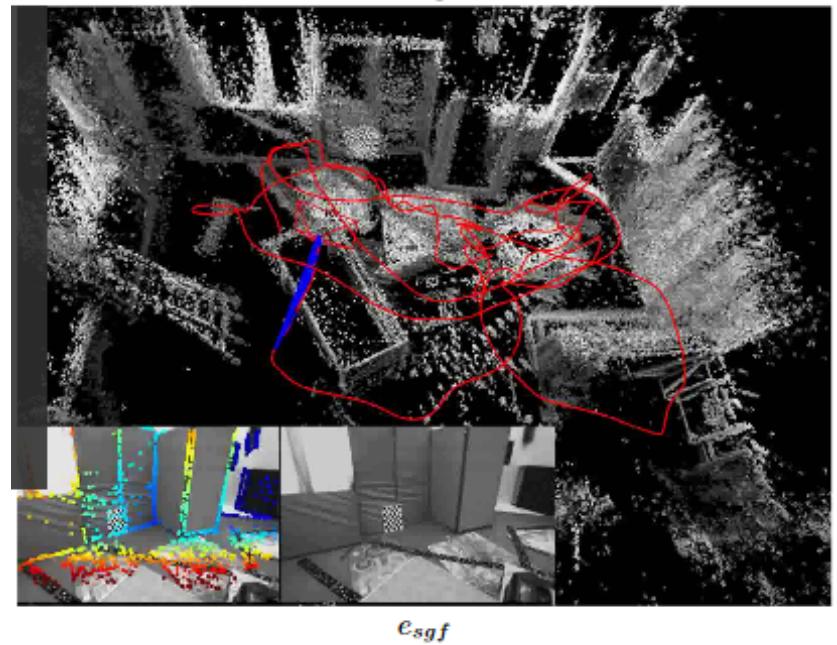
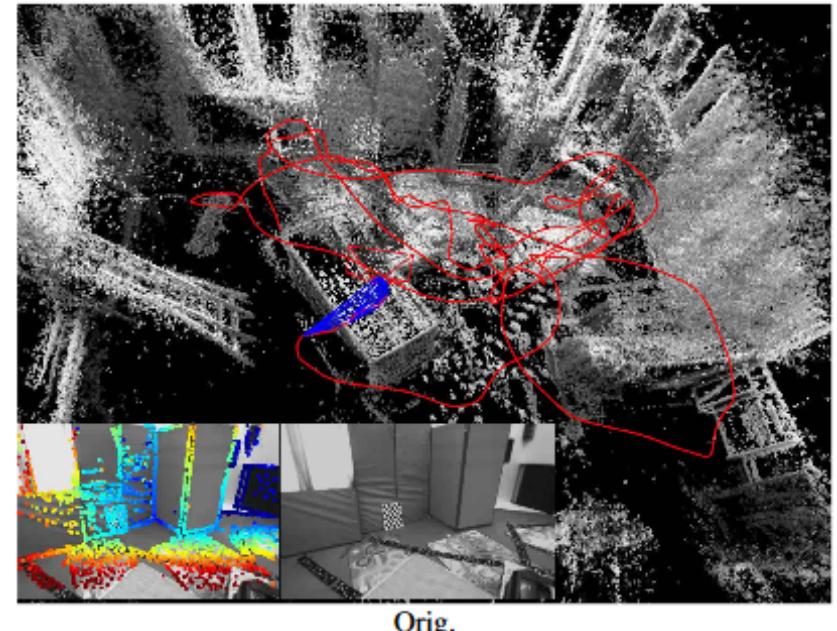
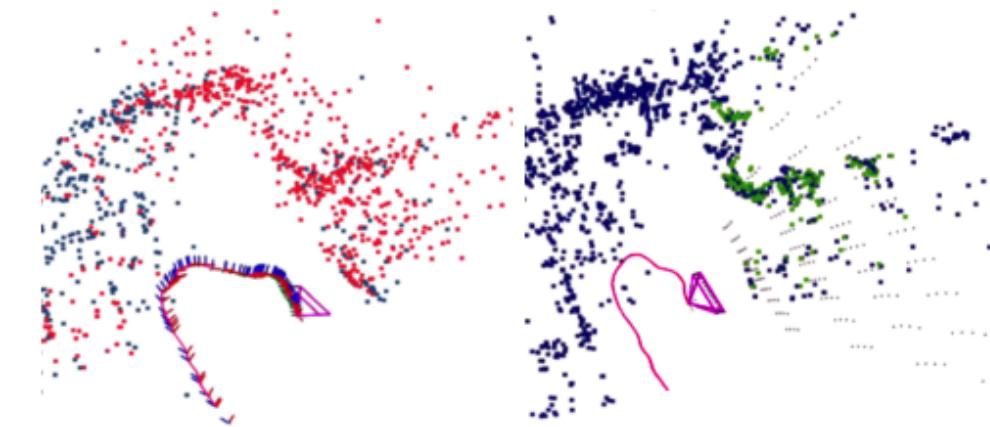


Fig. 6: Resulting map and trajectory (red line) of DSO [5] w/o and with  $c_{sgf}$  for depth estimation on V1\_01 of the EuRoC dataset [24]. The reduced drift is clearly visible in the sharper edges and an reduction of double walls.

## FutureMapping 2: Gaussian Belief Propagation for Spatial AI

Visual SLAMで用いられる従来のキーフレーム表現 (covisibility graph) では、3D点の遮蔽関係を記述できないため、ボクセルハッシングと視錐台表現を用いたレイキャスティングにより、高速かつ省メモリに幾何的関係性の記述を可能とした。

<https://arxiv.org/abs/2003.02247> pic.twitter.com/MMeABTUKVP



(a) Map query using keyframes. (b) Map query using voxel map.

Fig. 6: Results of querying keyframe-map and voxel-map at the same pose (purple pyramid) in a SLAM pipeline. Blue (both): all points in the map; Red: visible points queried from overlapping keyframes; Green: visible points queried from the proposed voxel-map; Gray: sample points in the raycasting-based query (Section II-C).