

## 深層学習ベースのStereo Visual SLAM手法（2021年以降）

近年（2021年以降）に発表された、ステレオカメラ入力に対応した深層学習ベースのVisual SLAM手法として、以下のようなオープンソース実装が存在します。それぞれ概要、精度、対応センサー、屋外環境への適性、開発状況（メンテナンス状況）などの点で比較します。

### DROID-SLAM (2021年) <small> – 深層学習による高精度SLAM</small>

【DROID-SLAM】はTeedらによって2021年に発表された深層学習SLAMシステムで、**モノクロ・ステレオ・RGB-D**いずれのカメラ入力にも対応しています<sup>①</sup>。特徴として、**畳み込みGRUベースの繰り返し最適化と差分可能な密束調整（Bundle Adjustment）**層を組み合わせ、カメラ姿勢と画素単位の深度を同時に高精度に推定します<sup>②</sup>。従来法に比べ**軌道誤差（ATE）の大幅低減**やロバスト性向上が報告されており<sup>③</sup>、リアルタイム動作しつつ全体的なドリフトを抑制しています。ステレオ（左右カメラ）やRGB-Dデータも**推論時に統合**できる柔軟性があり、モノクロで学習したモデルでもステレオ入力を加えることで精度と安定性を向上できるのが特徴です<sup>④</sup>。このため、自動運転車両やロボット、AR/VRなど**屋外・室内を問わず高精度な自己位置推定**が可能であり、長時間のトラッキングにおいても従来より頑健です<sup>⑤</sup><sup>③</sup>。オープンソース実装がGitHubで公開されており、2025年現在もマルチGPU対応の改善などコミュニティによるメンテナンスが継続しています（GitHub上で2000+stars、2025年5月にマルチGPU対応コミットあり）<sup>⑥</sup><sup>⑦</sup>。

- ・**概要・手法:** リカレントネットワーク+差分可能BAによるエンドツーエンドSLAM<sup>②</sup>。深度と姿勢を同時推定。
- ・**精度:** TartanAirやEuRoCなどで軌道誤差を**桁違いに削減**し、ドリフトや失敗を大幅に低減<sup>③</sup>。KITTI相当の野外シーケンスにも高精度。
- ・**対応センサー:** モノクロ / **ステレオ** / RGB-D（単一モデルで対応）。ステレオ入力時は**実スケール**での軌跡推定が可能<sup>④</sup>。
- ・**屋外環境適性:** シミュレーションから実環境まで**高い汎化性能**。車載カメラ等にも適用可能で、未学習のシナリオにもオンライン適応可能<sup>⑤</sup>。
- ・**開発状況:** **オープンソース（GitHub）**で公開<sup>①</sup>。継続的にアップデートされ、派生として軽量化したDPVO (2023)も公開されるなど活発です。

### Liuらの学習特徴量ベースStereo SLAM (2025年) <small> – CNN+GNNによるロバストSLAM</small>

Liuら（2025年）は、ORB-SLAM系のステレオSLAMパイプラインを**学習ベースの特徴検出・マッチング**に置き換えることで、難条件下でのロバスト性と精度を高めた手法を提案しています<sup>⑧</sup>。具体的には、**CNNによる特徴点抽出**および**GNN（グラフニューラルネット）による特徴マッチング**を組み込み、システム全体で**深層特徴量のみを地図表現に使用**します<sup>⑨</sup>。これにより大きな視点変化やテクスチャが乏しい場面でも従来のORB等より安定した対応付けが可能です。トラッキングやローカルマッピング段階で従来のヒューリスティックなマッチングを**学習ベース手法に置換**し、複雑な環境下での対応付け性能を向上させています<sup>⑩</sup>。さらに、**学習特徴量に基づくBag-of-Words辞書**を構築してループ検出・再ローカライゼーションにも利用しており、**ループクローズ**によるドリフト抑制も可能です<sup>⑩</sup>。公開データセット（おそらくKITTI屋外やEuRoC室内）および実機ロボットで評価した結果、**精度・頑健性で他の最新SLAMを上回る**と報告されていま

す<sup>11</sup>。特に**大きな並進・回転運動**や**低テクスチャ面**での追跡が改善し、屋外の広範囲環境においても高精度な軌跡推定が可能です。現時点で公式のコード公開情報は見当たりませんが、類似手法としては後述のMAC-VOやDH-PTAMでも学習特徴の活用がなされています。

- ・**概要・手法:** 従来の特徴ベースSLAMを**学習型特徴 (CNN+GNN)** に全面置換。深層特徴に統一した地図表現。
- ・**精度:** KITTIやEuRoCでORB-SLAM3など**従来法を上回る精度**。大きな視点変化やスパーステクスチャでも安定<sup>8 11</sup>。
- ・**対応センサー:** **ステレオカメラ** (左右同期画像)。※モノクロには非対応 (左右画像が深度推定に必須)。
- ・**屋外環境適性:** 高速な動きや遠景を含む屋外環境に強く、**モバイルロボット実験でも良好**<sup>11</sup>。ループ閉鎖も備え大規模環境対応。
- ・**開発状況:** 論文発表段階 (IEEE T. Instrum. Meas. 2025)。GitHubでの**公式実装公開は未確認**。学術成果としての位置づけで、今後の公開に期待。

## MAC-VO (2025年) <small> – 不確かさ学習によるStereo Visual Odometry</small>

MAC-VO (Metrics-Aware Covariance for VO) は2025年にQiuら (CMU) が発表しICRAベストペーパーを受賞した、**ステレオ視覚オドメトリ**の深層学習手法です<sup>12</sup>。特徴は**マッチングの不確実性を学習**し活用する点で、深層ネットワークにより画像間の対応点の**不確かさ (uncertainty)** を定量化します<sup>13</sup>。この不確かさ指標を用いて**信頼度の低い特徴点をフィルタ**し (キーポイント選択)、さらに**3次元点の共分散**をスケール考慮してモデル化することで、**ポーズグラフ最適化時に外れ値の影響を低減**しています<sup>13 14</sup>。従来の幾何VOがテクスチャエッジ重視の特徴を選んでいたのに対し、本手法ではグローバルな不整合に基づき**環境に応じた有効特徴のみを選択**するため、**照度変化やテクスチャの乏しい場面、動きの激しい場面でも高いロバスト性**を示します<sup>13</sup>。実際、公開データセット上で**最先端のVOアルゴリズムや一部のSLAMシステムをも凌ぐ精度**を達成し<sup>15</sup>、難条件下 (暗所や動的シーン) でも安定した自己位置推定が可能であると報告されています<sup>15 16</sup>。また各特徴点の推定精度を共分散として出力できるため、**姿勢結果の信頼度情報**も得られる点は自律システムの意思決定に有用です<sup>15</sup>。コードは公式GitHubに公開されており、ROS2パッケージも整備されるなど積極的にメンテナンスされています (2025年11月更新、800+stars)<sup>17 18</sup>。

- ・**概要・手法:** **不確実性学習**によるキーポイント選別と**メトリック共分散**の導入。学習した2次元マッチング不確かさから3D共分散を推定し後端最適化に組込<sup>14</sup>。
- ・**精度:** EuRoC (室内)、KITTI (屋外)、TartanAir等で**既存のVO/SLAMを上回る軌跡精度**。特に照明変化・動的物体・低テクスチャで高い追跡信頼性<sup>15</sup>。
- ・**対応センサー:** **ステレオカメラ専用** (左右同期画像によるVO)。深度センサ非使用・自己位置推定に特化。
- ・**屋外環境適性:** **屋内外を問わず頑健**。異なるスケールの環境 (室内vs屋外) でもスケール一貫性を保つ設計で、パラメータ調整なしで良好に機能<sup>19 14</sup>。
- ・**開発状況:** **オープンソース実装あり** (GitHub, MITライセンス)<sup>17</sup>。ROS2対応やドキュメントも充実しており、開発も活発 (ICRA 2025発表時点から継続更新中)。

## DH-PTAM (2024年) <small> – ステレオ映像＋イベントセンサの深層SLAM</small>

DH-PTAM (Deep Hybrid PTAM) はSolimanらによる手法で、**ステレオフレームカメラとステレオイベントカメラ**を統合した異種センサSLAMシステムです<sup>20</sup>。特徴は、**従来型PTAM** (Parallel Tracking and Mapping) の枠組みに**イベント駆動カメラ**の高速・高ダイナミックレンジな情報を組み込み、さらに**SuperPoint**など深層学習ベースの**特徴点検出・記述子**を用いてロバスト性を高めている点です<sup>21 22</sup>。ス

テレオ画像フレームとイベントストリームを**時空間的に同期**する新手法により、両者を共通の地図座標系で統合しつつ、イベントの高頻度情報で**高速運動やHDR環境**に対応します<sup>20</sup>。深層学習による特徴抽出・マッチングにより、暗所・激しい動きでも追跡性能が向上し、**大規模高照度差シナリオでも高い精度と頑健性**を示しました<sup>23</sup>。並列トラッキング・マッピングに加え**簡易なループ閉じ込み**も実装されており、長期運用でのドリフト抑制も図っています<sup>24</sup>。評価ではVECTorやTUM-VIEといったイベントデータセット（小規模から大規模屋外シーケンスを含む）で既存手法を上回る性能を確認しており、特に**夜間などHDR条件下でのロバスト性は従来比向上が顕著**です<sup>23</sup>。実装は研究目的のPythonコードですが**GitHubで公開**されており、イベント+ステレオの最先端SLAM手法として後続研究に供されています<sup>25</sup>。

- ・**概要・手法:** **ステレオ画像+ステレオイベントセンサ融合SLAM**。SuperPoint/HardNet/KeyNet等の学習特徴でフレーム・イベント双方の特徴検出<sup>22</sup>。PTAMベースの並列最適化+ループ閉じ込み。
- ・**精度:** 屋内外のイベントデータセットで**最先端の精度と頑健性**<sup>23</sup>。特に**高速運動や夜間環境**で他法より安定した自己位置推定を実現。
- ・**対応センサー:** **ステレオカメラ+ステレオイベントカメラ**（DAVISなどDVS）。フレームのみ/イベントのみでも動作可能だが、両方併用で性能向上。
- ・**屋外環境適性:** **高照度差（昼夜）環境や動きの激しい屋外シーンに強い**<sup>23</sup>。イベントセンサによりブレや露出オーバーに耐性があり、車両搭載にも有望。
- ・**開発状況:** **オープンソース実装あり**（GitHub公開, Python）<sup>25</sup>。Stars数は少ないものの2024年に論文掲載済みで、今後の研究発展に利用可能。

## DynaSLAM II (2021年) <small> – 動的物体を同時追跡する Stereo SLAM</small>

**DynaSLAM II**は、Mur-ArtalのORB-SLAMシステムを拡張した動的環境対応SLAMで、Bescosらによって2021年に提案されました。**ステレオカメラやRGB-Dカメラで動作し、インスタンスセマンティックセグメンテーション（Mask R-CNNなど）による物体検出とマルチオブジェクト追跡**をSLAMに統合しているのが特徴です<sup>26</sup>。具体的には、**動的物体を検出・追跡しつつORB特徴でカメラと物体の両方の軌跡を推定し、静的シーンの地図と動物体の構造・軌道を合わせて新たに提案されたバンドル調整で同時最適化**します<sup>27</sup>。各動体には3次元バウンディングボックスも推定して短時間ウィンドウで最適化しており、シーン中の動くエージェント（車や人）の**6DoF軌跡まで含めた地図生成が可能**です<sup>27</sup>。従来は動的物体を除去する手法が多かった中、本手法では**動物体も情報源とみなし積極的にトラッキング**することで、シーン理解に有用なだけでなく**カメラの自己位置推定精度も向上**することを示しました<sup>28</sup>。著者らの報告によれば、動的物体を同時に最適化することでカメラ軌道の誤差が減少し、人口密集環境など**従来不向きだったシナリオでも安定したSLAM**が実現できます<sup>28</sup>。オープンソースコードは初版DynaSLAM（2018年）のリポジトリ上で公開が予告されており<sup>28</sup>、既に先行研究者コミュニティで高い関心を集めています（※2025年現在、公開時期に関するIssueが立てられています）。DynaSLAM IIは特に**自動運転や複数ロボットの協調**など動的要素を無視できない屋外応用に適しており、Semantic SLAM分野のベースラインの一つとなっています<sup>29</sup><sup>30</sup>。

- ・**概要・手法:** **動的SLAM**。Mask R-CNNによる物体検出+Kalmanフィルタ等でMOT、ORB-SLAMベースのBAに**カメラと物体の姿勢を両方組込んで最適化**<sup>27</sup>。
- ・**精度:** **静止・動体双方の地図を同時に生成**することで、動的環境下でも高精度な自己位置推定を実現<sup>28</sup>。KITTI Trackingデータセットなどで静的SLAMとの差異を実証。
- ・**対応センサー:** **ステレオカメラ, RGB-Dカメラに対応**<sup>26</sup>（モノクロ版も理論上可能）。センサキャリブレーションによりメートルスケール軌跡を出力。
- ・**屋外環境適性:** **人や車の多い市街地シナリオに強い**。動的物体を除去せず活用するため、**交通シーンなどで安定したSLAM動作**<sup>29</sup><sup>30</sup>。静的地図+動的オブジェクト軌跡により環境理解も深い。
- ・**開発状況:** **DynaSLAM（初版）のコードは公開済み**<sup>31</sup>。DynaSLAM IIについても論文受理に合わせコード公開予定とされています<sup>28</sup>。Issueでも公開時期が問合せ中で、リリースされればORB-SLAM3派生の有力実装となる見込みです。

## 比較とまとめ

以上のように、深層学習を活用したStereo SLAMには多様なアプローチがあります。**特徴抽出・マッチング**の学習適用という点では、Liuらの手法やDH-PTAMがCNN特徴やGNNマッチング、SuperPointなどでロバスト性を高めています。一方、**最適化パイプラインへの学習統合**という点では、DROID-SLAMやMAC-VOがBAや重み付けにニューラルネットワークを組み込み高精度化を達成しています。また、**動的環境への対応**ではDynaSLAM IIがセマンティクスとSLAMの統合により他と一線を画します。いずれの手法も**ステレオカメラ入力からメートルスケールの3次元カメラ軌跡（6DoF姿勢系列）を出力**でき、高精度な自己位置推定を実現します。屋外の広域環境では、ORB-SLAM系伝統の全体BAやループ閉じ検出を備えるLiuら(2025)やDynaSLAM IIが**グローバルな整合性**で有利な一方、DROID-SLAMやMAC-VOは**局所的な精度・ロバスト性**で優れ、GPSなしでもドリフトの少ない軌跡が得られています。それぞれ公開実装も存在し（Liuら2025のみ未公開）、**開発の活発さ**ではDROID-SLAMやMAC-VOが現在も更新が続いている点で際立ちます。用途に応じて、たとえば**動的物体の多い屋外**ならDynaSLAM II、**暗所・高速移動**ならDH-PTAM、**汎用的な高精度VO**ならMAC-VOやDROID-SLAM、といった選択が考えられるでしょう。それぞれのGitHubリンクや論文を参考に、要件に適した手法を検討できます。

参考資料: DROID-SLAM <sup>2</sup> <sup>4</sup>、LiuらStereo SLAM <sup>8</sup> <sup>11</sup>、MAC-VO <sup>13</sup> <sup>15</sup>、DH-PTAM <sup>23</sup>、DynaSLAM II <sup>27</sup> <sup>28</sup>。

---

1 **GitHub - princeton-vl/DROID-SLAM**

<https://github.com/princeton-vl/DROID-SLAM>

2 3 4 5 **DROID-SLAM: A Differentiable Visual SLAM System**

<https://www.emergentmind.com/topics/droid-slam>

6 7 **Commits · princeton-vl/DROID-SLAM · GitHub**

<https://github.com/princeton-vl/DROID-SLAM/commits/main/>

8 9 10 11 **A Robust and Accurate Stereo SLAM Based on Learned Feature Extraction and Matching | Request PDF**

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/396550711_A_Robust_and_Accurate_Stereo_SLAM_Based_on_Learned_Feature_Extraction_and_Matching)

[396550711\\_A\\_Robust\\_and\\_Accurate\\_Stereo\\_SLAM\\_Based\\_on\\_Learned\\_Feature\\_Extraction\\_and\\_Matching](https://www.researchgate.net/publication/396550711_A_Robust_and_Accurate_Stereo_SLAM_Based_on_Learned_Feature_Extraction_and_Matching)

12 13 15 **MAC-VO: Metric-Aware Covariance for Learning-based Stereo Visual Odometry**

<https://mac-vo.github.io/>

14 16 19 **MAC-VO: Metrics-aware Covariance for Learning-based Stereo Visual Odometry mac-vo.github.io**

<https://arxiv.org/html/2409.09479v2>

17 18 **MAC-VO · GitHub**

<https://github.com/MAC-VO>

20 23 24 25 **[2306.01891] DH-PTAM: A Deep Hybrid Stereo Events-Frames Parallel Tracking And Mapping System**

<https://arxiv.org/abs/2306.01891>

21 **DH-PTAM: A Deep Hybrid Stereo Events-Frames Parallel Tracking ...**

[https://www.researchgate.net/publication/381368104\\_DH-PTAM\\_A\\_Deep\\_Hybrid\\_Stereo\\_Events-Frames\\_Parallel\\_Tracking\\_And\\_Mapping\\_System](https://www.researchgate.net/publication/381368104_DH-PTAM_A_Deep_Hybrid_Stereo_Events-Frames_Parallel_Tracking_And_Mapping_System)

22 **GitHub - AbanobSoliman/DH-PTAM: A Deep Hybrid Stereo Event-Visual Parallel Tracking And Mapping System**

<https://github.com/AbanobSoliman/DH-PTAM/>

26 27 28 29 30 31 **DynaSLAM II: Tightly-Coupled Multi-Object Tracking and SLAM | Request PDF**

[https://www.researchgate.net/publication/350361007\\_DynaSLAM\\_II\\_Tightly-Coupled\\_Multi-Object\\_Tracking\\_and\\_SLAM](https://www.researchgate.net/publication/350361007_DynaSLAM_II_Tightly-Coupled_Multi-Object_Tracking_and_SLAM)