# 単眼カメラ画像を用いた潜在的危険領域推定に関する研究

TP22013 下村 晃太　　指導教授：山下 隆義

## 1. はじめに

完全自動運転を実現するには，現時刻の周辺状況の理解だけでなく，未来に起こりうる危険につながるオブジェクトがどのように変化するかを予測することも重要である．特に，歩行者や自動車が突然現れる可能性の高い潜在的危険領域の予測は，安全な自動運転や交通事故の回避に不可欠である．自動運転支援の従来研究では，定義した危険領域の教師ラベルを学習済みモデルの予測結果を用いて生成する．生成した教師ラベルは，学習済みモデルの精度に依存するため，正確な潜在的危険領域の推定は困難である．

本研究では，歩行者や自動車が飛び出す恐れのある領域を潜在的危険領域と定義し，既存データセットに潜在的危険領域をアノテーションして拡張した新たなデータセットを作成する．また，作成したデータセットを用いた潜在的危険領域の推定手法を提案する．評価実験より，提案手法は様々なシーンで潜在的危険領域の推定が可能であることを示す．

## 2. 従来研究

Kozuka らは，歩行者が飛び出す恐れのある領域を危険領域と定義し，危険領域の中心 1 ピクセルにアノテーションを付与する手法と回帰ベースの損失関数を用いた危険領域推定手法を提案した [1]．Zhou らは，予測した深度マップの勾配情報を用いて死角を推定し，走行中の自車両に悪影響を与える歩行者を早期に検出する手法を提案した [2]．従来研究の定義する危険領域は，歩行者が突然飛び出す可能性がある領域や，カメラに映る自動車によって生じる死角に限定されている．また，Zhou らのように学習済みのモデルを用いた死角の推定手法は，死角に対する Ground Truth がないため，定量的な評価が困難である．

## 3. 提案手法

本研究では，Cityscapes データセット [3] に潜在的危険領域のアノテーションを付与した新たなデータセットを構築する．また，潜在的危険領域データセットを用いた潜在的危険領域推定手法を提案する．

### 3.1 潜在的危険領域データセットの構築

歩行者や自動車が突然現れる可能性の高い領域を，潜在的危険領域と定義し，人手でアノテーションを付与する．潜在的危険領域は，異なる物体間の境界である場合が多く，マスク等を用いた領域ベースのアノテーションは困難である．そこで，従来研究 [1] を参考に，Cityscapes データセットの潜在的危険領域の中心に対して 1 ポイントアノテーションを行った．ここでは，データセットを構成する都市ごとに自動車の運転経験を持つアノテーターを割り当て，各画像に対して最低一つの危険領域を選択することを条件とした．データセット全体にかかるアノテーション時間は 3385 分であり，1 枚あたり約 41 秒の時間を要した．図 1(b) にアノテーション例を示す．

(a) RGB画像　(b) 1ポイントアノテーション　(c) Ground Truth

図 1：提案手法で用いる潜在的危険領域のラベル例

### 3.2 距離情報を用いた Gound Truth の前処理

1 ポイントアノテーションによってピクセルレベルで表現される潜在的危険領域は，画像領域に占める割合が小さく，学習時に扱える情報量が少ない．また，画像内における距離情報を考慮した場合，遠方の潜在的危険領域は相対的に小さく，近い潜在的危険領域は相対的に大きくすべきである．そこで本研究では，距離情報をもとに潜在的危険領域の範囲を変化させることで前述の問題に対処する．具体的には，屋外画像の距離バイアス [4] と，Cityscapes が車載カメラ画像であることから考えられる画像内の消失点と，危険領域の座標間距離をもとにピクセルレベルのアノテーションを半径 $r$ の円状に拡大する．半径 $r$ の計算方法を式 (1) に示す．

$$r = \frac{|(x_r, y_r) - (x_v, y_v)|}{s} \tag{1}$$

ここで，$(x_r, y_r)$ は任意の潜在的危険領域の座標，$(x_v, y_v)$ は消失点の座標，$s$ はスケール係数である．拡大した領域内にガウシアンフィルタを適用することで，潜在的危険領域の連続的な空間変化を表現する．学習に利用する Ground Truth の例を図 1(c) に示す．

### 3.3 潜在的危険領域推定ネットワーク

従来研究 [1] に倣い，提案手法のネットワークにはエンコーダ・デコーダ構造を採用する．図 2 に提案手法のネットワーク構造を示す．マルチスケール特徴を抽出するために，異なるカーネルサイズを持つ複数の畳み込み層を組み合わせた Inception モジュールを用いる．Inception モジュールは，受け取ったテンソルに対して，異なるカーネルサイズの畳み込み処理を行い，チャンネル次元に沿って結合する．提案する潜在的危険領域推定ネットワークは，複数の Inception モジュールと残差接続から構成する．同色のブロックは，共通の Inception モジュールである．また，$\oplus$ は要素ごとの加算を示す．

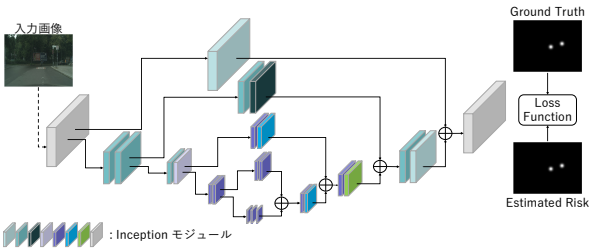


図 2：ネットワーク構造

### 3.4 損失関数

本タスクにおける学習目標は，様々なシーンにおいて画像内に存在する局所的な潜在的危険領域を正確に推定することである．そこで本研究では，新たに Exponentially Weighted MSE Loss と Total Variation Distance から構成する Relative Risk Loss を提案する．

**Exponentially Weighted MSE Loss**. 車載カメラ画像に対する潜在的危険領域は，一般的な Saliency Prediction タスクと比べ，画像領域に対する Ground Truth の示す領域の割合はわずかである．このような不均衡な問題を解決するために，Exponentially Weighted MSE Loss を導入する．式 (2) に示すように，MSE (Mean Squared Error) に対して，予測値の大きさに応じて指数関数で重み付けをする．これにより，従来の MSE に比べて疎な Ground Truth に対する結果として 0 を予測する傾向に対応する．

$$\mathcal{L}_{Ew-MSE} = \frac{1}{N}\sum_{i=1} exp(-z_i)(I_i - z_i)^2 \tag{2}$$

ここで，$I \in \mathbb{R}^{(1*h*w)}$ は Ground Truth，$z \in \mathbb{R}^{(1*h*w)}$ はモデルの出力，$N \in \mathbb{R}^{(h*w)}$ はピクセル数である．

**Total Variation Distance Loss**. 潜在的危険領域を推定する場合，ピクセル単位で回帰・分類するために設計された損失関数を用いると，各ピクセル間の関係を考慮することができない．そこで，潜在的危険領域を任意の画像における潜在的危険領域の確率分布とみなし，確率分布間距離を最小化する損失関数を導入することで，前述の問題に

対処する．ここでは，従来研究 [5] に倣い線形正規化を用いて確率分布に変換する．Total Variation Distance Loss を以下に示す．

$$\mathcal{L}_{Tv-Dist} = \sum_i |f(z_i) - f(I_i)| \tag{3}$$

$$f(z_i) = \frac{x_i^z}{\sum_{i=1}^N x_i^z}, \qquad f(I_i) = \frac{x_i^I}{\sum_{i=1}^N x_i^I} \tag{4}$$

ここで，$x = (x_1, ..., x_N)$ は推定されたリスク確率 ($x^z$)，または，Ground Truth($x^I$) に対する正規化されていない値の集合である．

**Relative Risk Loss** 最終的な Relative Risk Loss は，Exponentially Weighted MSE Loss と Total Variation Distance Loss から構成される．潜在的危険領域推定ネットワークは，Relative Risk Loss を用いて最適化する．

$$\mathcal{L}_{Risk} = \lambda \mathcal{L}_{Ew-MSE} + \mathcal{L}_{Tv-Dist} \tag{5}$$

ここで，$\lambda$ は Exponentially Weighted MSE Loss の係数である．

## 4. 評価実験

未知画像に対する潜在的危険領域の推定精度により提案手法の有効性を検証する．従来手法の Point supervision [1] と比較する．統一的な比較実験のため，ネットワークは共通とした．事前学習時は，Cityscapes セグメンテーションデータ，ファインチューニング時は，潜在的危険領域データを用いる．そのため，事前学習時は，最終層をセマンティックセグメンテーション用のヘッドに置き換えて学習する．ファインチューニング時は，最終層を潜在的危険領域推定ヘッドに置き換え，それ以外は事前学習済みの重みを用いて学習する．

### 4.1 定量的評価

表 1 に潜在的危険領域推定の定量的評価を示す．先行研究 [5] に倣い，Area Under Curve (AUC) と Correlation Coefficient (CC) を用いて評価する．表 1 より，提案手法は，従来手法と比較して，全ての Cityscapes データセットにおいて，AUC と CC の両方で優れた精度を達成した．このことから，Rilative Risk Loss を組み込むことで，不均衡な潜在的危険領域の Ground Truth に対応することができ，全体的にパフォーマンスが改善したといえる．特に，提案手法は，従来手法と比較して，テストデータに対する CC のスコアが 0.118pt 向上しており，潜在的危険領域の優先度を考慮した推定が可能である．また，セマンティックセグメンテーションタスクによる事前学習済みモデルを用いた場合，検証データ，テストデータの両方で精度の向上が確認できた．この結果から，潜在的危険領域推定タスクにおいて，事前学習タスクとしてセマンティックセグメンテーションを用いることが有効であることを確認した．

表 1：Cityscapes での定量的評価

| 比較手法 | 事前学習 | Cityscapes | | | |
|---|---|---|---|---|---|
| | | 検証データ | | テストデータ | |
| | | AUC | CC | AUC | CC |
| Point supervision [1] | | 0.545 | 0.111 | 0.544 | 0.105 |
| | ✓ | 0.513 | 0.097 | 0.512 | 0.098 |
| 提案手法 | | 0.540 | 0.158 | 0.542 | 0.183 |
| | ✓ | **0.562** | **0.218** | **0.564** | **0.216** |

### 4.2 定性的評価

提案手法と従来手法の潜在的危険領域の推定結果を可視化することで，提案手法の有効性を調査する．定性的評価を図 3 に示す．1 行目のシーンでは，提案手法は，右側の停車車両の間を潜在的危険領域と推定している．このとき，遠方の車両周辺よりも近距離の車両周辺に対して，危険度が高いと推定している．それだけでなく，対向車に対しても危険と推定している．その一方で，従来手法は，潜在的危険領域として推定可能な範囲が狭く，また遠方の車両に対しては推定ができていない．2 行目のシーンでは，遠方の歩行者と比べて自車両との距離が近い歩行者に対して，危険度が高いと推定している．提案手法は，車両だけでなく歩行者に対する潜在的危険領域が推定可能であることが分かる．3 行目のシーンでは，提案手法は，自車両を右後ろ側から追い抜く車両と遠方車両に対して潜在的危険領域を推定している．このとき，近距離の車両の方が危険度が高くなっている．一方で，従来手法は，距離に基づく危険度の変化を考慮することができていない．4 行目は T 字路のシーンであり，提案手法は，カメラの画角外につながる道路から飛び出す可能性を考慮した潜在的危険領域の推定結果となっている．その一方で，従来手法は，画像内の車両や歩行者を潜在的危険領域と推定する傾向があり，画面外から飛び出してくる可能性を考慮した潜在的危険領域の推定ができていない．従来手法は，事前に予測したセグメンテーションマスクを用いて教師ラベルを新たに生成している．教師ラベルの生成時，道路は安全な領域と定義するため道路上の潜在的危険領域に対する推定能力が低下したと考えられる．以上から，提案手法は，シーン全体のコンテキストを捉えた潜在的危険領域推定が可能であることが分かる．

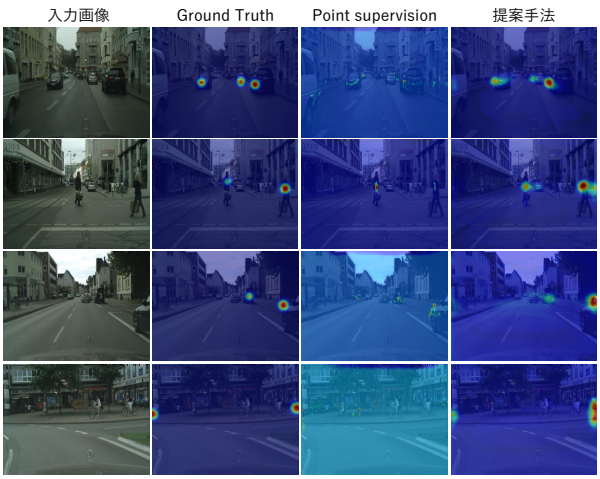


図 3：定性的評価結果

## 5. おわりに

本研究では，潜在的危険領域を推定するための新たなデータセットとベースライン手法を提案した．提案手法は，潜在的危険領域推定タスクを確率分布予測タスクとして定式化し，線形正規化ベースの損失関数を採用した．実験から提案手法は従来手法と比較して優れた性能を示すことを確認した．今後は，提案手法の有効性を調査するために，KITTI 等の大規模データセットでの実験を行う．

## 参考文献


[1] K.Kozuka, *et al.*, “Risky Region Localization With Point Supervision”, In IEEE International Conference on Computer Vision (ICCV) Workshops, 2017.

[2] J.Zhou, *et al.*, “High-Speed Recognition of Pedestrians out of Blind Spot with Pre-detection of Potentially Dangerous Regions”, IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2022.

[3] M.Cordts, *et al.*, “The Cityscapes Dataset for Semantic Urban Scene Understanding”, In IEEE IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[4] W.Chen, *et al.*, “Single-Image Depth Perception in the Wild”, Advances in Neural Information Processing Systems, 2016.

[5] S.Yang, *et al.*, “A dilated inception network for visual saliency prediction”, In IEEE Transactions on Multimedia, 2019.


## 研究業績


[1] K.Shimomura, *et al.*, “Potential Risk Estimation with Single Monocular Camera”, In Secure and Safe Autonomous Driving Workshop and Challenge on CVPR, 2023.

（他 2 件）