

畳込みニューラルネットワークを用いた熊検出及び精度評価

徳竹 正行[†] 富岡 洋一[†] 齋藤 寛[†] 小平 行秀[†]

会津大学 コンピュータ理工学部[†]

1. はじめに

近年，地方で熊の目撃情報が増加しており，人々の安全を守るために熊を迅速に発見し注意喚起することが求められている．そこで我々は画像認識を用いて熊を自動検出するセンサーノードの開発を目指している．

本稿では，熊候補領域抽出と畳込みニューラルネットワークによる画像分類を組み合わせて，画像中に熊が存在するか否かを判定する熊検出手法を提案する．また，ツキノワグマの画像を用いて，熊の検出精度を評価し，熊の大きさやシーンの明るさ，隠蔽の有無が熊検出精度に及ぼす影響について報告する．

2. 熊検出手法

熊検出センサーノードではカメラで撮影した静止画像を熊検出器に入力し，画像内に熊が存在するか否かを判定する．

図1に熊検出器の処理の流れを示す．まず，入力画像から熊が存在する可能性の高い熊候補領域を抽出する．次に，それぞれの熊候補領域を画像分類用の畳込みニューラルネットワークに入力し，熊かそれ以外かを判定する．少なくともひとつの熊候補領域が熊と判定されたならば，入力画像中に熊が存在すると判定する．

本稿では，画像分類用の畳込みニューラルネットワークとしてImageNet Large Scale Visual Recognition Competitionの1000クラスの画像分類用データセットに対する学習済モデルであるInception v3 [1]を利用した．また，各熊候補領域に対して，Inception v3の推論結果の上位5位以内にツキノワグマ，ヒグマ，ナマケグマのいずれかが含まれるとき，その候補領域は熊であると判定した．

3. 熊候補領域抽出手法

物体の候補領域の抽出手法としては，Selective search [2]が提案されているが，この手法は入力画像内のあらゆる物体を囲む矩形領域を抽出しようとするため計算量が多い．一方，本研究では熊領域のみ抽出することが目的である．また，評価実験により，熊候補領域中に熊

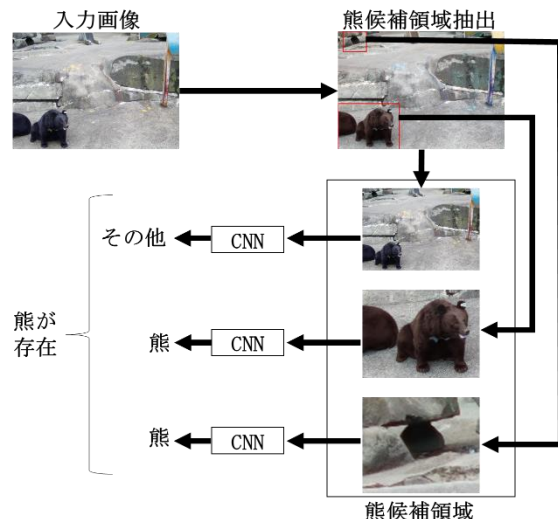


図1 熊検出器の処理の流れ．図中のCNNは畳込みニューラルネットワークを表す．

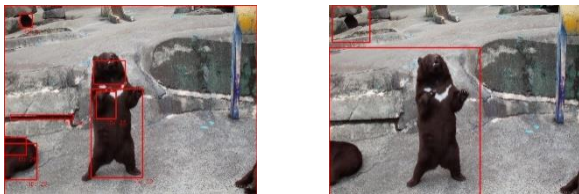
以外の物体が含まれているか，熊が画像中の一部にしか映っていないかとも比較的高い確率でInception v3の推論結果の上位5位に熊が含まれることを確認した．このため，提案手法ではより単純な熊候補領域の抽出手法を導入する．

提案手法では，熊の体の色特徴を利用して，熊に対応する画素を抽出する．各画素の色特徴として，RGB値とHSV値を用いる．まず，入力画像中でRGB値かHSV値の少なくともひとつが指定範囲に含まれる画素を1，それ以外を0とする2値画像を生成する．熊に対応する画素の抽出例を図2に示す．次に，ラベリング処理により8近傍の連結成分に分解し，各連結成分の矩形領域を求める．抽出漏れを考慮し，各矩形領域を周囲 s 画素だけ拡大する．最後に，共通部分を持つ2つの矩形領域を，それらを囲む最小の矩形領域に統合する操作を繰り返すことで領域数を削減する(図3)．統合により残った領域と入力画像全体を合わせて熊候補領域とする．実験では，抽出する画素値の範囲を $R \leq 128, G \leq 40, B \leq 40, H \leq 140, S \leq 130, V \leq 50$ とした．但し， H は0から179，それ以外の値は0から255の範囲で正規化されているものとする．また， $s = 30$ とした．



(a) 入力画像 (b) 抽出結果

図 2 熊に対応する画素の抽出例



(a) 統合前 (b) 統合後

図 3 候補領域の統合

4. 実験結果

ツキノワグマを対象とし、検出精度の評価を行った。入力画像サイズは、640×480とした。機械学習フレームワークとして TensorFlow を使用した。

熊候補領域として撮影画像全体のみを考える基本方式と提案の熊候補領域抽出手法により抽出した領域を用いる改良方式の2方式について評価を行った。

北秋田市が運営するくまぐま園で撮影した 889 枚のツキノワグマの画像を Positive クラスとし、Caltech Pedestrian Detection Benchmark [3]の動画からランダムに抽出した 10000 枚のツキノワグマが含まれない画像を Negative クラスとして、適合率(Precision)、再現率(Recall)、F 値(F-measure)により評価した。再現率は熊検出率に対応する。実験結果を表 1 に示す。改良方式を用いることで基本方式に対して再現率が 4.3%向上し、F 値は 2.2%向上した。

次に IMAGENET データベースやインターネットより収集した、様々な条件下で撮影されたツキノワグマの画像 591 枚を用いて精度評価を行った。これらの画像を目視により確認し、ツキノワグマの撮影距離を3段階(遠距離, 中距離, 近距離)に、シーンの明るさを2段階(明るい, 暗い)に、

表 1 二方式の精度比較

| | 適合率 | 再現率 (熊検出率) | F 値 |
|------|--------|---------------|-------|
| 基本方式 | 100.0% | 93.5% | 96.6% |
| 改良方式 | 99.8% | 97.8% | 98.8% |

隠蔽の有無で2種類に分類しラベル付けを行った。

表 2 に各カテゴリの画像群に対する再現率を示す。改良方式を用いることで遠距離や暗いシーンの画像群に対して、特に精度を向上できていることがわかる。一方、熊のサイズが特に小さい場合や、シーンが暗い場合、木や岩などで熊の体が約半分以上隠れる場合には検出が困難な傾向があった。これらの特徴を有する画像に対する精度向上は今後の課題である。

5. 結論

熊候補領域の抽出と画像分類を組み合わせた熊検出手法を提案し、97.8%の熊検出率を達成した。

遠距離で撮影した場合、暗いシーン、隠蔽のある場合の熊検出率向上のため、学習モデルを改良することが今後の課題である。また、今後は本手法をセンサーノードに実装し、フィールドテストを実施する予定である。

謝辞

本研究は、総務省戦略的情報通信研究開発推進事業(SCOPE)の支援の下、行われたものである(課題番号 181602001)。

参考文献

- [1] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," arXiv:1512.00567, 2015.
- [2] Koen E. A. van de Sande et al., "Segmentation As Selective Search for Object Recognition," ICCV, 2011.
- [3] P. Dollár, C. Wojek, B. Schiele and P. Perona. "Pedestrian Detection: A Benchmark," CVPR, 2009.

表 2 様々な条件下での再現率(熊検出率)の比較

| | 撮影距離 | | | シーンの明るさ | | 隠蔽 | |
|------|-------|-------|--------|---------|-------|-------|-------|
| | 遠 | 中 | 近 | 暗 | 明 | 有 | 無 |
| 基本方式 | 90.8% | 98.4% | 99.2% | 90.0% | 97.1% | 94.3% | 97.8% |
| 改良方式 | 92.9% | 98.8% | 100.0% | 93.4% | 97.9% | 95.5% | 98.6% |