

深層学習を用いた新生児熱画像の部位検出

別府 文香¹ 吉川 寛樹¹ 内山 彰¹ 東野 輝夫¹ 濱田 啓介^{2,3} 平川 英司⁴

概要：新生児は体温調節機能が未熟なため、保育器を適切な温度に管理することが不可欠である。現在は、プローブで測定した体表温度を基に、医療者が手動で保育器の温度管理を行っている。しかし、新生児の皮膚は未熟なため、プローブが剥がれやすく、長時間の安定した測定は難しい。これに対して、サーモグラフィを用いることで、新生児に違和感やストレスを与えることなく体表温度の測定が可能となる。一方、サーモグラフィを用いた場合には、得られた熱画像から部位ごとの体表温度を抽出する必要がある。そこで本研究では、深層学習を用いて、新生児の熱画像から頭部・胸部・四肢の6箇所を検出する手法を提案する。提案手法では、まずYOLOv5を用いて身体部位6箇所の検出モデルを構築する。さらに、決定木に基づき部位間の相対位置を考慮したフィルタリングを適用する。性能評価のため、新生児26名の熱画像4820枚を使用した。その結果、適合率、再現率はそれぞれ94.8%、77.5%となった。また、検出した部位を基に首元の温度を取得し、プローブで測定した食道温度との相関を調べた結果、相関係数0.82となり、体温取得方法として有効であることが分かった。

1. はじめに

新生児は体温調節機能が未熟であり、外気の影響を受けて体温が変動しやすい。このため、保育器を適切な温度に保つことが重要である[1]。多くの医療現場では、プローブで測定した体表温度及び保育器内温度を基に、医療者が手動で保育器温度を調節している。しかし、プローブによる測定は新生児にとって侵襲性が高い。また、新生児の皮膚が未熟であることに加えて、動きが活発なため、皮膚に貼り付けたプローブが脱落しやすく、継続して安定した体温を測定することが難しい。

非侵襲的で信頼性の高い皮膚温測定のために、近年、サーモグラフィが注目されている。サーモグラフィは保育器内に設置するため、新生児に違和感やストレスを与えず、動きがある場合でも長時間の安全な測定が可能である。一方、サーモグラフィを用いた場合には、熱画像中のどの部分がどの身体部位に該当するかを特定する必要がある。これによって、測定したい部位の体表温度を抽出できる。身体部位の特定には、カメラを使用する方法も考えられるが、システムが複雑になり、導入やメンテナンスのコストが高くなる。また、新生児でもプライバシーの問題が生じる可能性がある。以上の理由から、熱画像のみから身体部位を検出することが望ましい。

熱画像においては、夜間における監視などの目的で、顔を対象とした認識手法が多数提案されている[2]。また、顔以外も含めた身体部位を検出する試みとして、OpenPose[3,4]を熱画像用に改変したThermalPose[5]が存在する。しかし、ThermalPoseは複数人の姿勢推定を目的としており、比較的骨格が明確に判別できる成人を対象としているため、そのまま新生児に適用することはできない。一方、可視画像における身体部位検出では、OpenPoseなどの姿勢推定手法が多く提案されている。多くの手法は機械学習を利用しており、熱画像は可視画像と異なるため、熱画像用に学習をし直す必要がある。このためには、熱画像にラベル付けを行い、新たなデータセットを構築する必要がある。その労力は大きいので、適切な手法の設計が重要である。

そこで本研究では、保育器内の新生児熱画像から、深層学習を用いて6つの身体部位を検出する手法を提案する。提案手法では、YOLOv5[6]^{*1}を基に、頭部、胸部、左右の upper 及び lower の6つの部位を検出するモデルを作成する。これら6部位は、首元などの詳細な部位を見つける手がかりとなる主要な部位として選定した。YOLOv5では部位間の位置関係を考慮しないため、提案手法では決定木を利用して、検出部位の整合性を確認することで性能向上を図る。

性能評価のため、新生児26名から合計4820枚の熱画像を収集した。その結果、身体部位検出の適合率は94.8%、再現率は77.5%となった。また、提案手法により抽出した体表温度の有用性を示すため、提案手法の検出結果に基づ

¹ 大阪大学大学院情報科学研究科

² 長崎みなとメディカルセンター臨床工学部

³ 長崎大学大学院医歯薬学総合研究科

⁴ 鹿児島市立病院新生児内科

^{*1} <https://github.com/ultralytics/yolov5>



図 1: データ収集環境

き首元の体表温度を抽出し、食道温との相関を調べた。その結果、相関係数は 0.82 となり、新生児の非侵襲的な体温取得方法として有効であることが分かった。

2. データ収集

データ収集環境を図 1 に示す。新生児の全身が熱画像内に収まるよう、サーモグラフィは保育器上部に設置した。熱画像のサイズは横 320×縦 256 ピクセルである。最終的に、26 名の新生児から熱画像 4820 枚を取得した。

収集した熱画像には、医療者による介入中の状態や、器具の映り込みが存在する。また、新生児が横を向いているなど、対象身体部位の一部が映っていない場合が存在する。このような画像を使用すると、正しい部位検出を行うことが難しい。そのため本研究ではそのような画像を無効画像と定義し、正しく部位検出を行いやすい有効画像との分類を事前に行う。分類方法の詳細は 3.2 節で述べる。最終的に、3868 枚の熱画像が無効、952 枚が有効と判断され、有効画像のみに身体部位のラベル付けを行った。

3. 提案手法

3.1 手法概要

手法の概要を図 2 に示す。まず、医療者による介入が行われている状態や、体勢によって身体部位の判別が難しい場合の画像を取り除くため、CNN による二値分類を行う。次に、有効画像と分類された画像に対して、YOLOv5 を用いて構築した部位検出モデルにより部位検出を行う。検出部位は、頭部・胸部・左右の上肢及び下肢の 6 箇所である。さらに、検出された部位に対し、頭部を基準とした相対位置の特徴量を抽出し、決定木に入力する。決定木により、相対位置を考慮したうえで、入力された部位がどの部位に該当するかが出力される。この決定木の出力部位と、YOLOv5 により検出された部位が一致する場合のみ、正しい結果とみなす。最後に、検出した各部位の位置に基づき、所望する部位の体表温度を取得する。本稿では一例として、頭部と胴体の検出位置から首元の体温を抽出する方法を述べる。

3.2 熱画像分類

身体部位検出に有効画像のみを使用するため、CNN による二値分類を行う。有効画像と無効画像には、形や大きさに加えて温度分布などの違いがあり、比較的容易に判別が可能である。それに合わせてネットワークを小さくすることで、不要なパラメータの影響を抑え、過学習を防いでいる。初期学習率を決める手段として、学習率の範囲を決定する方法としてよく用いられている LR range test [7] を用いた。LR range test は、一定幅で学習率を徐々に増加させて変化を調べる方法であり、本研究では最も損失が低い学習率 0.00001 を採用した。入力画像に対して、 3×3 のフィルタによる畳み込み処理を 1 層目では 16 回、2 層目では 32 回行うネットワークを構築した。学習率は 0.00001、バッチサイズは 16、エポックは 100 とし、活性化関数として ReLU を用いて学習を行う。

3.3 身体部位検出

身体部位の検出には、YOLOv5 により構築した身体部位検出モデルを用いる。部位検出モデルは、CNN で分類された有効画像を入力とし、矩形とクラス名（部位名）が出力される。ネットワークは、パラメータ数 700 万個の YOLOv5s モデルを使用し、クラス数は、身体部位数である 6 に設定した。バッチサイズは 16、エポックは 1500 に設定して学習を行った。

想定環境において、保育器内に存在する新生児は 1 人のみであり、同一部位が複数存在することはない。したがって、提案手法では各部位について YOLOv5 が出力した結果のうち、最も信頼度が高い結果のみを採用する。図 3 に検出結果の例を示す。

3.4 部位判別モデルによるフィルタリング

部位検出モデルにおいては、部位間の位置関係を考慮していない。このため、部位の位置関係として存在しえない場合がある。このような誤った検出結果を除去するため、部位間の相対位置や大きさなどを特徴量として、決定木により部位判別モデルを構築する。

部位判別モデルの入力は、部位検出モデルで出力された各部位の特徴量である。決定木は、これらの特徴量を入力として、該当する部位を出力する。

部位検出モデルによって部位 p として出力された矩形を B_p とする。また、部位判別モデルに対して、矩形 B を入力として与えた時の判別部位を $D(B)$ とする。提案手法では、 $p = D(B_p)$ の時、 B_p を有効な結果として採用し、 $p \neq D(B_p)$ の時、 B_p を誤った結果として棄却する。

表 1 に示した通り、各検出結果の相対距離、方向ベクトルを含む、6 個の特徴量を用意した。相対位置を算出する際には、頭部を基準としている。この理由は、部位検出において頭部の精度が最も高く、安定して出力されるためである。決

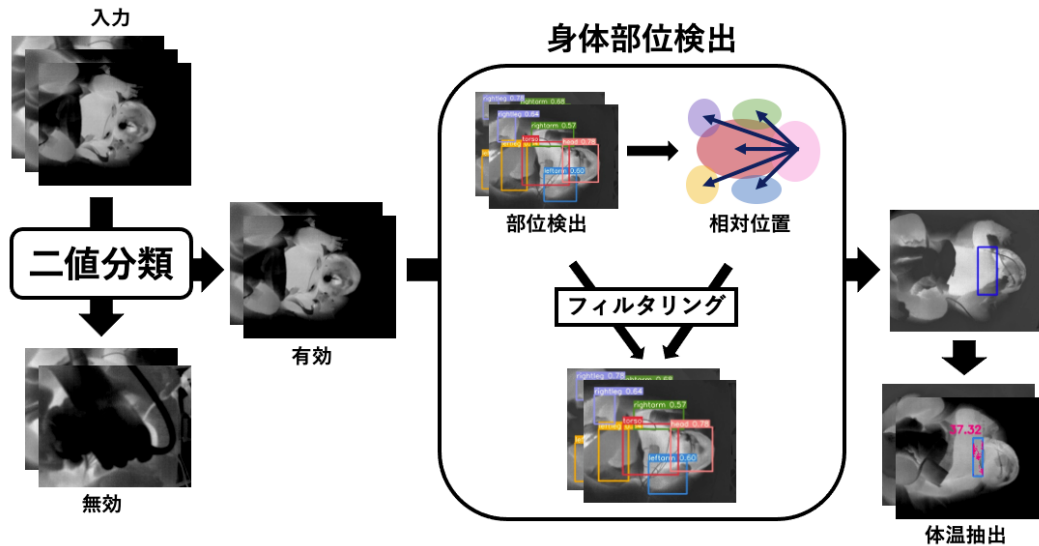


図 2: 提案手法の概要

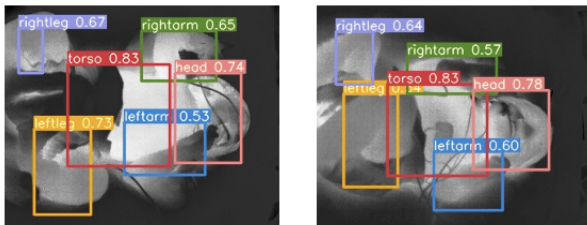


図 3: 部位検出結果の例

表 1: 部位判別モデルの特徴量

型	特徴量
絶対位置	x 座標, y 座標
大きさ	矩形面積
相対位置	頭部との距離, 頭部から x,y 方向へのベクトル

定木の実装には、scikit-learn の DecisionTreeClassifier^{*2}を利用した。

3.5 首元温度の抽出

提案手法を用いた体表温度測定の有効性を確認するため、首元温度を抽出する手法を設計した。サーモグラフィの位置は固定されているため、新生児の頭部は常に熱画像の右側に位置する。そのため、首元は頭部と胸部の間であると仮定する。

具体的には、首元の中心座標 c_n を以下のように定義する。

$$c_n = \frac{l_h + r_t}{2}$$

ここで、 l_h と r_t は、それぞれ頭部矩形の左辺の中点と胸部矩形の右辺の中点を示す。首の領域は、 c_n を中心として、高さ 60 ピクセル、幅 20 ピクセルを選択した。

次に、首元の矩形から体表温度を抽出する。この矩形に

は背景を含む他の領域を含んでいる可能性があるため、首元の矩形の温度分布のうち、上位 25% の平均値を取得する。首元温度は他の部位と比較して高いため、このような方法で首元温度を取得可能である。

4. 性能評価

4.1 評価環境

性能評価では、2 章で説明したデータセットを使用した。データセットは 3868 枚の無効画像と 952 枚の有効画像の計 4820 枚で構成されている。CNN の学習には、有効画像と無効画像からランダムに 762 枚ずつを選び、合計 1524 枚を用いた。部位検出モデルと部位判別モデルの学習を行う際にも、CNN の学習に用いた 762 枚の有効画像を使用し、残りの 4058 枚の熱画像をテストに用いた。身体部位のラベルは有効画像に対してしか与えられていないため、部位検出と部位判別の評価は有効画像のみを使用して行った。一方で、体表温度抽出の評価においては、学習データ以外の無効画像も含む全ての熱画像を用いて評価を行った。

評価指標として、物体検出や分類問題で一般に使われる適合率と再現率を用いる。さらに物体検出では、画像分類と異なり、クラス分類に加えて検出領域の正しさを評価する必要がある。このため、検出領域の大きさと位置の両方を考慮した Intersection over Union (IoU) [8] を用いた評価を行う。

$$\text{IoU} = \frac{|B_p \cap \hat{B}|}{|B_p \cup \hat{B}|},$$

ここで、 B_p はラベル付けされた身体部位 p に対する予測矩形、 $\hat{B}$ は真値矩形である。予測矩形と真値矩形が近ければ近いほど、IoU は高い値となる。

身体部位の大まかな位置を正しく検出することができれば、温度分布を基に体表温度を抽出することが可能である。

^{*2} <https://scikit-learn.org/>

表 2: 画像分類の結果

		予測		
		有効	無効	再現率 (%)
真値	有効	176	14	92.6
	無効	273	3595	92.9
	適合率 (%)	39.2	99.6	—

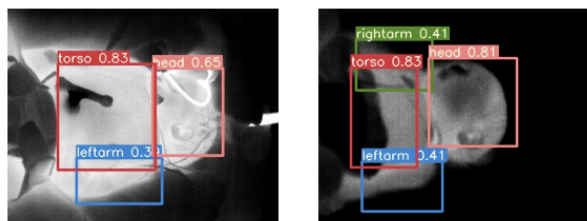


図 4: 誤分類された無効画像に対する部位検出結果

表 3: 部位検出の結果

	適合率	適合率 (フィルタ)	再現率	再現率 (フィルタ)
頭部	94.8	94.8	96.4	96.4
胸部	95.8	96.4	94.7	94.7
右上肢	87.7	94.9	67.5	63.7
左上肢	81.2	92.0	66.3	65.9
右下肢	92.7	95.8	67.5	66.9
左下肢	87.6	93.5	79.3	77.4
全体	89.9	94.8	78.6	77.5

そのため、以降の評価では、言及が無い限り IoU が 0.3 以上の場合を正しい検出結果とする。

4.2 画像分類の性能

熱画像に対して有効、無効の二値分類を行った結果を表 2 に示す。適合率は 69.4%、再現率は 92.8% となった。無効画像 3868 枚、有効画像 190 枚という不均衡なデータを使用しているため、有効熱画像の適合率は特に低いが、再現率についてはどちらも 92% を超えた。

誤って有効画像を無効と分類した場合は有効画像の損失となるが、誤って無効画像を有効と分類した場合は、身体部位の誤検出につながる可能性がある。一方、評価の結果、誤って有効と分類された無効画像は、有効画像と非常に良く似ていることが分かった。例えば、下肢のみが写っていないなど、存在しない部位があるような熱画像がこれにあたる。このような熱画像であっても、提案手法では、部位検出により画像に写っている部位のみを検出できる。図 4 に、無効であるにも関わらず有効画像と分類された場合の身体部位検出の例を示す。このように、身体部位検出の性能低下につながる無効画像の除去という点で、画像分類の性能は十分であると考えられる。

4.3 部位検出の性能

表 3 に部位検出の結果を示す。部位検出モデルのみを用

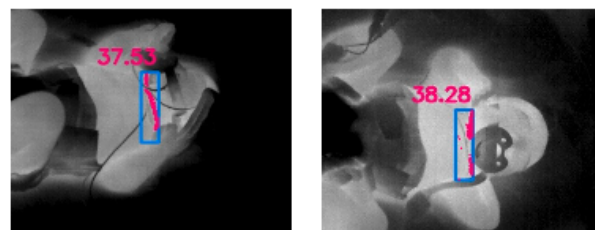


図 5: 首領域 (矩形) と温度分布上位 25% (矩形内の赤い点群) の例

いた場合、適合率は 89.9%、再現率は 78.6% であった。適合率が 95% 近くある頭部や胸部と比較して、四肢は 90% 前後と小さい。この原因は、四肢は頭部や胸部と比較して部位の大きさが小さく、動きが多いためだと考えられる。

これに対して、部位判別モデルによるフィルタリングを行った場合の適合率は 94.8%、再現率は 77.5% であった。部位検出モデルによって適合率は 4.9% 向上し、全ての部位において適合率は 90% を超えた。一方で、フィルタリング前と比較して、再現率は 1.1% 低下している。これは、フィルタリングによって、一部の正しい結果を誤って除外してしまったことを示している。医療応用を想定する場合、たとえ再現率が低くとも、適合率が高いことが重要と考えられる。したがって、部位判別モデルによるフィルタリングを適用することは有効である。

4.4 体表温度の抽出結果

図 5 に、検出結果を基に決定した首領域の矩形と抽出温度の例を示す。温度分布の上位 25% のピクセルのほとんどが首周りに集中していることが分かる。抽出した首元の体表温度の有用性を確認するため、取得温度とプローブで測定された食道温度との相関を調べた。性能評価では、無効画像と有効画像の両方を使用しているため、誤って有効と分類された無効画像からも、首元温度が抽出される可能性がある。しかし、無効画像であっても、頭部と胴体が検出されていれば、首元温度を抽出することは可能である。したがって、全ての部位が検出されなくとも、頭部と胴体が検出されていれば、首元温度を抽出した。最終的に、228 枚の熱画像から首元温度の抽出に成功し、そのうち有効画像は 170 枚、無効画像は 58 枚であった。これらのうち、同時に食道温の測定が行われていた 26 枚の熱画像を用いて、食道温度と抽出した首元温度の相関を調べた。

図 6 に、首元温度と食道温度の相関を示す。相関係数は 0.82 となり、両者に強い正の相関が見られた。この結果は抽出された体表温度が中枢温度の基準として使用できることを示しており、本手法の有用性を確認できた。

5. おわりに

本研究では、深層学習を用いた熱画像における新生児

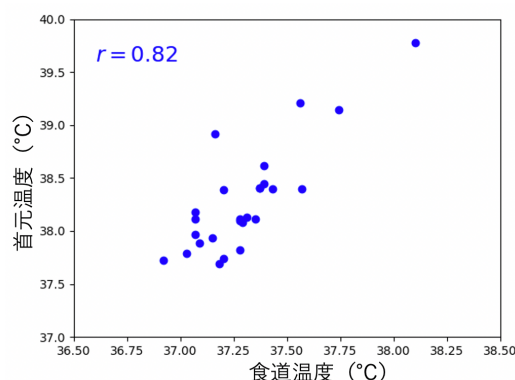


図 6: 食道温度と抽出した首元温度の相関

の身体部位検出モデルを提案した。提案手法では、CNN、YOLOv5、決定木を組み合わせることで性能向上を図った。新生児 26 名の熱画像 4820 枚を使用して評価を行った結果、部位検出において適合率 94.8%、再現率 77.5%を達成した。さらに、検出した部位を基に首元の体表温度を取得し、食道温度との関係を調べたところ、相関係数 0.82 の強い正の相関があることを確認した。

今後の課題として、さらにデータを増やして、食道温度と首元の体表温度の関連性を分析することが挙げられる。また、今回は首元温度のみを取得したが、検出した四肢を利用して、抹消温度の取得も行う予定である。さらに、体表温度分布や保育器内温度と深部体温の関連調査を引き続き進める予定である。

参考文献

- [1] Knobel, R. B.: Thermal Stability of the Premature Infant in Neonatal Intensive Care, *Newborn and Infant Nursing Reviews*, Vol. 14, No. 2, pp. 72–76 (2014).
- [2] Krišto, M. and Ivašić-Kos, M.: An overview of thermal face recognition methods, *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1098–1103 (2018).
- [3] Cao, Z., Hidalgo Martinez, G., Simon, T., Wei, S. and Sheikh, Y. A.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 43, No. 1, pp. 172–186 (2019).
- [4] Cao, Z., Simon, T., Wei, S.-E. and Sheikh, Y.: Real-time multi-person 2D pose estimation using part affinity fields, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1302–1310 (2017).
- [5] Chen, I.-C., Wang, C.-J., Wen, C.-K. and Tzou, S. J.: Multi-Person Pose Estimation Using Thermal Images, *IEEE Access*, pp. 174964–174971 (2020).
- [6] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A.: You Only Look Once: Unified, Real-Time Object Detection, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788 (2016).
- [7] Smith, L. N.: Cyclical Learning Rates for Training Neural Networks, *Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472 (2017).
- [8] Rezaatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A.,

Reid, I. and Savarese, S.: Generalized Intersection over Union, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666 (2019).