

深層学習を用いた新生児熱画像の部位検出に基づく 体温抽出手法の検討

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、CNNを用いて対象部位が見えない画像を分類した上で、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、検出部位を基にk-means法を用いたクラスタリングを適用し、末梢温度や首元温度を取得する。新生児26名の熱画像4820枚を使用して性能評価を行った結果、部位検出の適合率、再現率はそれぞれ94.8%、77.5%となった。また、取得温度とプローブで測定した体温には強い正の相関が見られ、体温取得方法として有効であることが分かった。末梢温度の抽出においては、クラスタリングを適用することで背景を除外し、末梢部位の温度分布を抽出可能な見通しを得た。

キーワード: 深層学習, 部位検出, 新生児, 熱画像

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か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できることが望ましい。

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。その労力は大きいため、適切な手法の設計が重要である。

そこで我々の研究グループでは、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している [6]。提案手法では、YOLOv5 [7]^{*1} を基に、

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

頭部、胸部、左右の上肢及び下肢の 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。

本稿では、この部位検出法に基づき、首元や末梢温度を抽出する手法の検討を行う。部位検出モデルでは 6 つの身体部位を検出可能だが、実際に体温を取得したい部位としては、手足や指先などの末梢部や首元などが想定される。この場合、部位検出モデルでは該当する部位そのものを検出できるわけではないため、そのままでは体温抽出に利用できない。

体温を抽出したい部位そのものを検出する方法も考えられるが、形状に特徴が少ないため、その実現は容易でない。このため、本研究では検出が比較的容易な 6 部位の位置関係を利用して、首元や末梢の体温抽出を試みる。首元は頭部と胸部のほぼ中間に安定して位置するため、検出した頭部と胸部の中間点を固定幅で拡張することでおよその首元領域を決定する。この首元領域の温度分布に対して、上位 25% を首元温度として抽出する。この理由は、首元温度は深部体温と相関があることが知られており [8]、他部位の体表温よりも温度が高い傾向を示すためである。一方、末梢部位は検出矩形の大きさが検出結果によって異なるため、同様の方法を適用できない。このため、手足の部位検出結果に対して k-means 法を用いたクラスタリングを適用し、新生児と背景を分割した上で温度を抽出する。

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、抽出した首元温度と食道温との相関を調べた結果、相関係数は 0.82 となり、強い正の相関が見られた。さらに、末梢温度の抽出においては、クラスタリングを適用することで背景を除外し、末梢部位の温度分布を抽出可能な見通しを得た。

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、頭部と胴体の検出位置から首元の体温を抽出する方法と、四肢から末梢の体温を抽出する方法の検討を行う。

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 [9] 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 のフィルター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



図 2: 提案手法の概要

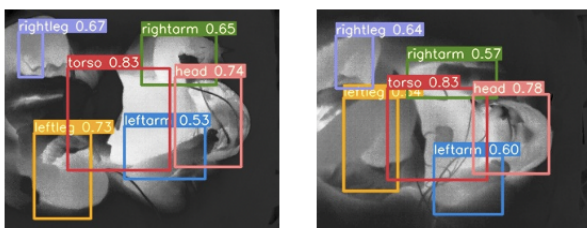


図 3: 部位検出結果の例

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

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

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 midpoint と胸部矩形の右辺の midpoint 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。

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

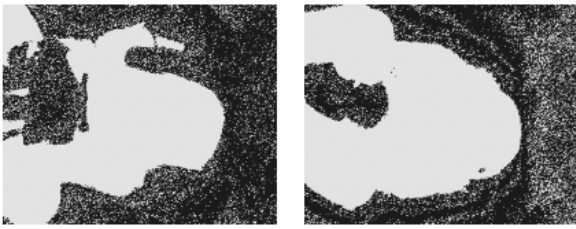


図 4: クラスタリング結果の例

これは、首元温度は深部体温と相関があることが知られており、他の部位と比較して高い傾向を示すためである。

3.6 末梢温度の抽出

首元温度の抽出では、首元領域を固定の大きさを持つ矩形として定義している。一方、末梢温度の抽出では、上肢・下肢の検出矩形に対して、矩形内の体表温度分布を抽出する必要がある。矩形の大きさは検出結果によって変化するため、矩形内に含まれる体表領域の割合も変化する。このため、首元温度の抽出と同様に温度分布の上位一定割合を体表温度として抽出する手法は適切でない。そこで、我々はクラスタリングを用いて温度分布の類似部分を抽出し、背景と新生児の分離を試みる。本稿では、クラスタリング手法として軽量の k-means 法を適用することとした。

$$E = \sum_{x_j \in X} \min_{i \in k} \|x_j - c_i\|^2$$

ここで、 $X_i, j \in 1, \dots, n$ は総数 n のデータ、 c_i はクラスタの中心である。新生児とそれ以外を分離するため、クラスタ数 k は 2 に設定する。k-means 法では、 c_i から各データ $x_i, j \in 1, \dots, n$ の距離の総和、つまり評価関数 E が最小となるような c_i を、クラスタに変化がなくなるまで繰り返し求める。これにより、類似した温度を持つピクセルが同一クラスタに統合され、結果として背景と新生児が分割される。図 4 に、クラスタリング結果の例を示す。

末梢温度の抽出では、クラスタリングを用いて高温/低温クラスタの 2 種類に分割を行い、部位検出結果の矩形内の高温クラスタの平均値を当該部位の温度とする。これにより、背景のように低温な体表以外の領域を除外し、体表の温度分布を取得できる。

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、

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	-

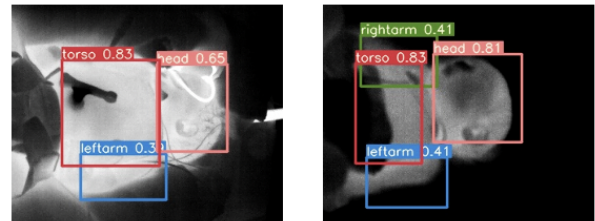


図 5: 誤分類された無効画像に対する部位検出結果

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) [10] 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。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

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

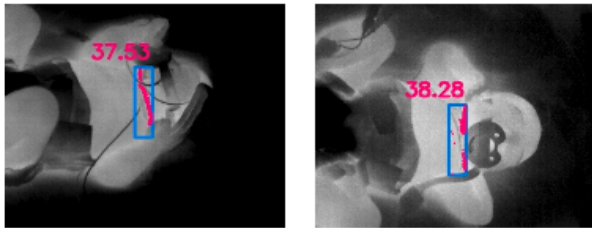


図 6: 首領域 (矩形) と温度分布上位 25% (矩形内の赤い点群) の例

た。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 5 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 首元温度の抽出結果

図 6 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

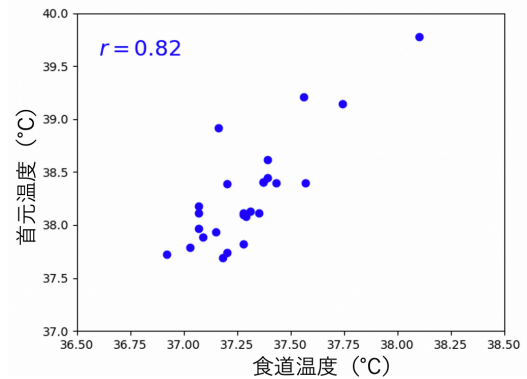


図 7: 食道温度と抽出した首元温度の相関

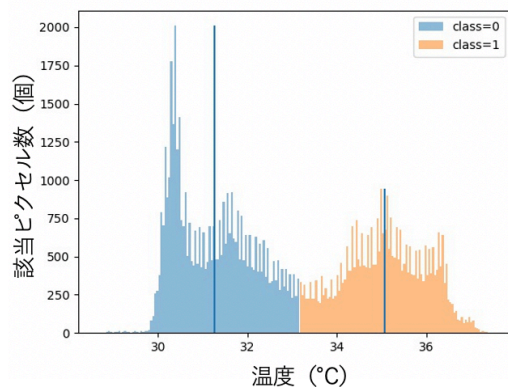


図 8: 温度分布とクラスタリング結果

表 4: 末梢温度抽出の結果

	矩形内平均	上位 25% 平均	クラスタ平均
右上肢	35.46	37.06	35.51
左上肢	36.08	36.87	35.79
右下肢	36.12	36.59	35.92
左下肢	34.66	36.31	35.39
末梢温度差	1.46	0.75	0.53

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 7 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。

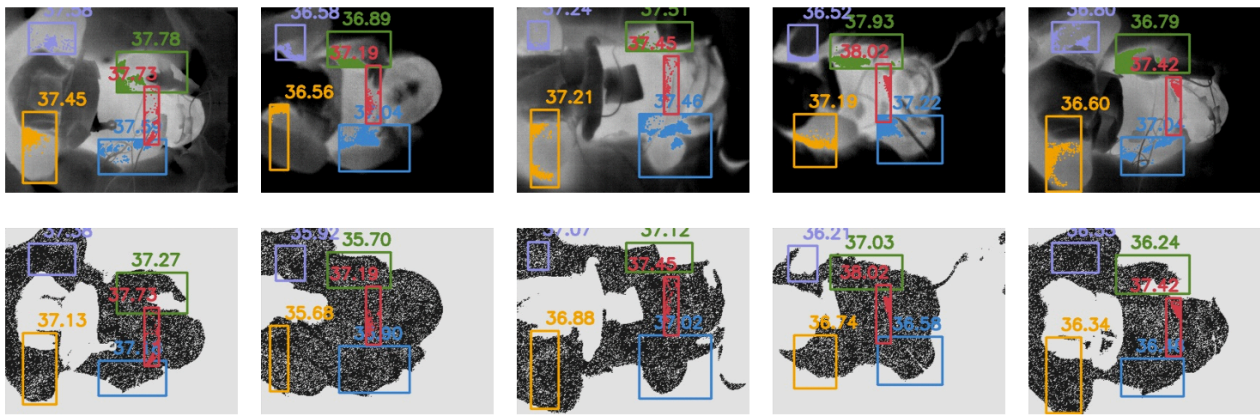


図 9: 温度抽出結果の画像例. 上段の色付き部は各部位における温度分布の上位 25%を示しており, 数値はその平均値を示している. 下段は提案手法により特定した高温クラスターの平均値を示す.

4.5 末梢温度の抽出結果

図 8 に, 各クラスターの温度分布を示す. class=0 に分類されたクラスターは温度が低く, class=1 に分類されたクラスターは温度が高いため, class=1 は新生児の体表部である可能性が高い.

部位検出結果を基に, 末梢温度を取得した画像例を図 9 に示す. 図の上段, 下段はそれぞれ, 検出矩形内の温度分布上位 25%の平均値を取得した場合, および検出矩形内の高温クラスターの温度分布平均値を取得した場合である. この結果より, 上位 25%を抽出した場合, 上肢の高温部分は手の付け根付近や脇が多く, 下肢の場合は膝裏に集中していることが見てとれる. また, これらの箇所は中枢に近いため, 首元温度と近い傾向がある. このため, 末梢温度を抽出する方法としては相応しくないと考えられる. 一方, クラスタリングを適用した場合には, 検出矩形内の体表と背景を分離できており, 対象部位の温度分布抽出に有効なことが示唆される.

また, 1枚の熱画像に対して, 抽出した末梢温度を表 4 に示す. 表 4 では, 検出矩形に対して, 平均温度, 温度分布の上位 25%平均, 高温クラスターの平均を示している. 矩形内の平均温度を取得した場合, 末梢温度の差は最大で 1.46 度となっており, 最も大きい. また, 左下肢の温度が 34.66 度と, 他の箇所と比較して非常に低い. これは, 検出矩形内に背景として映り込む毛布など, 新生児以外の箇所の温度も含まれていることが原因であると考えられる. 一方, 上位 25%の平均値を取得した場合, 全体的に取得温度が高いことが分かる. これは, 図 9 から分かるように手の付け根付近や脇, 膝裏などの特定部位の温度が取得されているためである. これらに対して, クラスタリングを適用した場合, 部位間の温度差は 0.53 度と最も低く, ばらつきが小さい. 以上の結果より, クラスタリングを適用することで, 背景を除外し, 末梢部位の温度分布が適切に抽出可能な見通しを得た.

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. そして, 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. さらに, 末梢温度の抽出において, k-means 法を用いたクラスタリングを適用し, 新生児と背景を分離した上で温度を抽出する手法を検討した. 評価結果より, クラスタリングを適用することで適切に背景と体表部を分離可能であり, 末梢部位の温度分布抽出に有効な見通しを得た.

今後の課題としては, 厳密なクラスタリング評価を行うことが挙げられる. 現段階では, ピクセルレベルでの部位ラベル付けを行っていないため, 取得温度や可視化結果を用いた考察にとどまっている. このため, 今後部位毎のラベル付けを行い, 温度分布の比較などによって定量的な評価を行う予定である. また, 取得した複数部位の温度を基に医学的に重要な深部体温の推定法を設計することも重要と考えている.

参考文献

- [1] Robin B. Knobel. Thermal stability of the premature infant in neonatal intensive care. *Newborn and Infant Nursing Reviews*, Vol. 14, No. 2, pp. 72–76, 2014.
- [2] Mate Kristo and Marina Ivašić-Kos. An overview of thermal face recognition methods. In *International Conference on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1098–1103, 2018.
- [3] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 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] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1302–1310, 2017.
 - [5] I-Chien Chen, Chang-Jen Wang, Chao-Kai Wen, and Shiow-Jyu Tzou. Multi-person pose estimation using thermal images. Vol. 8, pp. 174964–174971, 2020.
 - [6] 別府文香, 吉川寛樹, 内山彰, 東野輝夫, 濱田啓介, 平川英司. 深層学習を用いた熱画像における新生児の身体部位検出法の検討. 情報処理学会研究報告, Vol. 2021-MBL-98, No. 2, pp. 1-2, 2021.
 - [7] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.
 - [8] 遠藤芳子, 武田淳子, 大池真樹, 丸山真紀子. 電子体温計による腋窩体温と前額部深部温との比較検討. 宮城大学看護学部紀要, 2009.
 - [9] Leslie N. Smith. Cyclical learning rates for training neural networks. In *Proc. of IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 464–472, 2017.
 - [10] Hamid Rezaatofghi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666, 2019.