

U-Netによるサーモグラフィデータの 特定位置予測技術に関する研究

Study on specific location prediction techniques for thermography data using U-Net

豊吉巧也¹⁾, 山路啓太²⁾, 瀧千波³⁾

Takuya Toyoshi, Keita Yamaji and Chinami Taki

1) 博(工) 愛知工科大学 工学部 機械システム工学科 准教授

(〒443-0047 愛知県蒲郡市西迫町馬乗50-2, E-mail: takuya.toyoshi@g.aut.ac.jp)

2) 愛知工科大学 工学部 機械システム工学科 学生

(〒443-0047 愛知県蒲郡市西迫町馬乗50-2, E-mail: 20203037@g.aut.ac.jp)

3) 摂南大学 学生部 スポーツ振興センター 助教

(〒572-8508 大阪府寝屋川市池田中町17-8, E-mail: chinami.taki@setsunan.ac.jp)

Face detection technology using RGB image data is widely used in many fields through computer vision and machine learning technologies. Thermography data is commonly used in conjunction with RGB images to acquire information on specific positions, and there are few examples of it being used alone. In this study, we examine technology to predict specific positions using only thermography data and propose a framework for learning and prediction. Specifically, we learn and predict the positions of the face, neck, jaw, and brow in thermography data that shows a human face. At the same time, we also examined a method for converting information obtained from images and actual temperature information and used U-Net to learn and predict the temperature field. We consider the learning process from the learning curves and describe the evaluation of the predicted results.

Key Words : Thermography data, U-Net, Specific positions prediction

1. はじめに

サーモグラフィカメラは、計測対象の温度分布を非接触で計測および記録可能な装置である。この装置は、コロナウィルス感染拡大防止のための、体温計測をきっかけに建物の入口や医療機関の受付などの多くの場所で目にする機会が増えた。これに加えて、近年の温暖化により発生の機会の増加が懸念される、運動時の熱中症予防や対策への活用が期待される装置でもある。

このような背景から、近年のサーモグラフィカメラは一般的なRGB画像も同時に撮影可能なデバイスも普及し、RGB画像を併用した活用が進んでいる。このようなデバイスから取得したデータを用いて、機械学習や深層学習によるデータの活用事例[1]には、RGB画像を併用したものが多く、サーモグラフィデータのための活用例は少ない[2]。この理由としては、サーモグラフィデータの解像度が低い点や、RGB画像を温度情報に対する付加情報として利用している点などが考えられる。

しかし、サーモグラフィデータのみでの物体検知やその活用法が広がることで、製造工程の簡略化や機器の開発および製造コストの低下につながり、RGB画像を必要としない場面での活用が進むことが期待できる。

そこで本稿では、深層学習技術のひとつであるU-Net[3]を用いてサーモグラフィデータの温度情報のみによる特

定部位の検出と温度場の学習モデルの構築手法について検討する。具体的な対象として、熱中症予防のための顔検出および特定部位の検出とその温度場の予測に関する学習モデルの構築手法について検討した。

2. サーモグラフィデータと学習データについて

HIKMICRO社製のPocket1を使用し、幅144 pixel、高さ192 pixelのサーモグラフィデータを取得した。本稿で使用した機種では、サーモグラフィデータは図1(a)のようなJpeg形式で取得することができる。このデータ内より得られる温度情報に対して、図1(b,c,d)のように顔、首、額のそれぞれの位置にラベル付けを行なった。ラベル付けはすべて楕円で行い、顔と首のラベルデータの重なり合う部分を顎のデータとし、首のラベルデータは元のラベルデータから顔との重なりあう部分を取り除いたものを用いた。これにより、顔全体が収められている1つのサーモグラフィデータから、顔、首、顎、額の4つのラベルデータを作成可能となる。また、図1(a)に示すサーモグラフィデータには、画像中央および画像内の最大、最小の温度が記載されており、この情報と整数で得られる画像情報から各pixelの温度の変換が可能となる。

このような手順で、合計159人分のサーモグラフィデータを取得し、ラベルデータの作成を行なった。本研究の実施にあたっては摂南大学「人を対象とする研究倫理審査

委員会」の承認を得て実施した（2024-042）．すべてのサーモグラフィデータに記載されている画像中央および最大，最小の温度と該当するpixelのデータをそれぞれ線で繋いだものを図2に示す．図2より，実際の温度と画像から取得できる値は線形変換で表されることが確認できる．しかし，計測時の環境により数値で表現される範囲が，1枚毎に異なることがわかった．これらの点を考慮し，学習モデルは特定部位検出および温度予測の2種類の学習モデルを生成することとした．

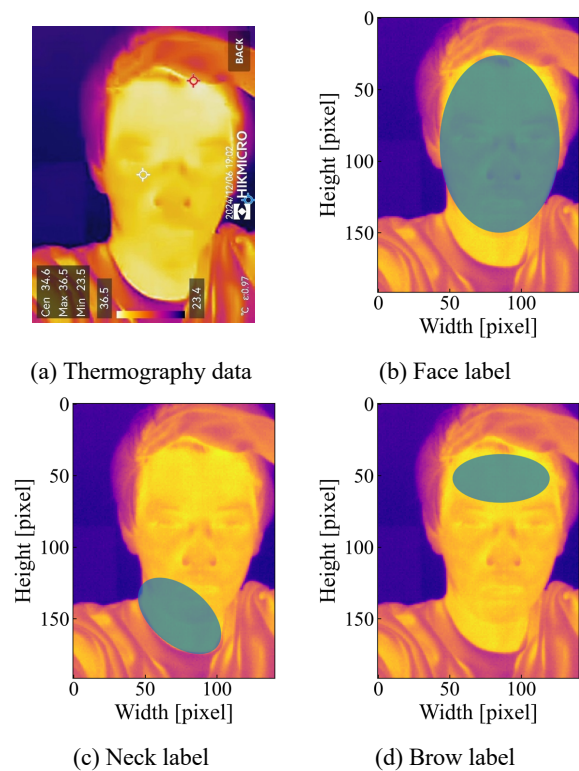


図-1 サーモグラフィデータとラベルの作成例

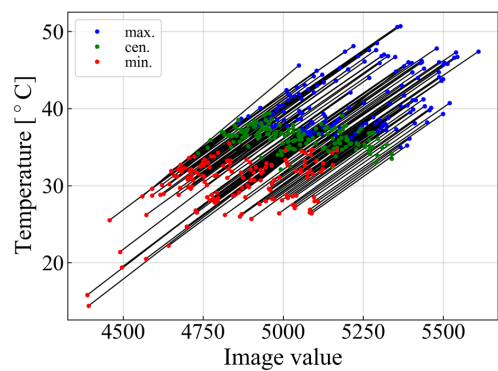


図-2 温度情報とサーモグラフィデータとの関係

3. 学習モデルについて

学習モデルには図3に示すU-Netを採用した．U-Netは，セマンティックセグメンテーション用のモデルであり，生体医学分野にて提案された学習モデルのひとつである．

U-Netは図3の左半分のエンコーダ部と右半分のデコー

ダー部からなる．図3右下に記載しているとおり，各矢印は各レイヤー間での畳み込みやプーリングの処理を示しており，入力から出力までの過程がアルファベットのUの字に似ていることからU-Netと呼ばれている．

図3に示すエンコーダ部とデコーダ部にかけて示される矢印は，U-Netにおける特徴のひとつであるスキップ接続(skip connection)である．これは，エンコーダ部での特徴マップ情報をデコーダ部に伝える役割を果たし，各pixelでの分類，予測精度の向上が期待できる処理である．

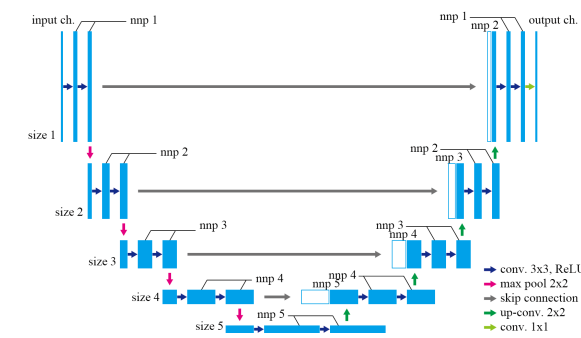


図-3 実装したU-Net モデル

学習モデルは，顔の特定部位検出用と温度予測用はそれぞれmodel 1,model 2として同時に学習を行う．ここで，model 1の入力データのチャンネル数(input ch.)はおよび，出力データのチャンネル数(output ch.)はそれぞれ，4とし，model 2の入力データのチャンネル数および，出力データのチャンネル数は，ともに1とした．学習モデルごとの各畳み込み層でのニューラルネットワークのパラメータ数(nnp)は表1に示す値を用いた．

表-1 各層のニューラルネットワークのパラメータ数

Number of neural network parameters (nnp)	Face parts detection (model 1)	Temperature prediction (model 2)
1	32	64
2	64	128
3	128	256
4	256	512
5	512	1024

また，図3のプーリングにより変化するデータのサイズ(size 1～5)は表2に示す値を用いた．各層での活性化関数にはReLU関数を用い，出力時においてmodel 1ではReLU関数，model 2ではSigmoid関数を用いた．最適化関数はAdamを用いて学習率はmodel 1では 1×10^{-3} ，model 2では 1×10^{-4} を用いた．そして，各層のニューラルネットワークパラメータの初期値には，xavierの初期値を用いた．

表-2 各プーリング層におけるデータサイズ

size	height	width
1	192	144
2	96	72
3	48	36
4	24	18
5	12	9

学習において、159データのうち30データを検証用として用いた。また、バッチサイズ $n = 10$ 、ラベルの正解値を y 、ラベルの予測結果を \hat{y} として損失関数を式 (1) のように定義し、エポック数を1000として学習を行なった。学習時の損失値に対して、各モデルの重みと正則化パラメータを用いて、式 (2) のように正則化を行なった。

$$\mathcal{L} = \frac{1}{n} \sum_i^n (\hat{y}_i - y_i)^2 \tag{1}$$

$$\begin{aligned} \mathcal{L} &= \mathcal{L} + \alpha \|w\|^2 \\ \text{where } \alpha_{\text{model 1}} &= 1.0 \times 10^{-3}, \\ \alpha_{\text{model 2}} &= 1.0 \times 10^{-4} \end{aligned} \tag{2}$$

4. 結果および考察

(1) 学習モデル生成時の学習曲線

学習モデル(model 1, model 2)の学習曲線を図4に示す。学習用データに対する損失値 (\mathcal{L}) は、エポック数が増加するにしたがい、両モデル共に低下していることが確認できる。model 1では、検証用データに対してエポック数600付近までは低下していることが確認できるが、それ以降では損失値がやや上昇し、過学習が生じていることが確認できる。一方のmodel 2においては、損失値は安定して低下した。model 1において、損失値が上昇した要因として、顔の特定部位に関するラベル情報は学習データに大きく依存する点と、顔、首、顎、額の4種類を同時に学習する点から、検証用データに対する汎化性能の低下が生じたものと考えられる。

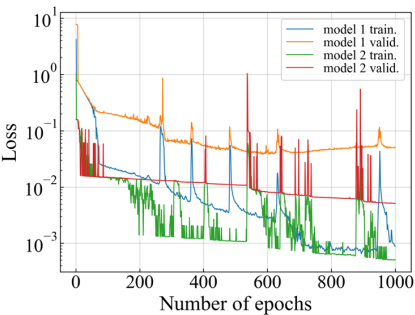


図-4 各学習モデルの学習曲線

(2) 顔の特定部位の位置と温度場の予測

学習後の各学習モデル(model 1, model 2)による予測結果の一例を図5に示す。図5(a.1)はサーモグラフィデータを正規化した入力データ、図5(a.2)は温度場の予測結果を示している。また、図5(b.1-4)には顔、首、顎、額の順で各部位の箇所の予測結果を示している。

顔の特定部位については、未学習のデータに対しても顔、首、顎のおおまかな位置の予測は可能であることが示されたが、額については背景の一部も同時に検出している結果を示した。また、温度場については、何が映っているかの判断がつく程度の予測は可能ではあるものの、背景の温度が高く予測されている結果を示した。

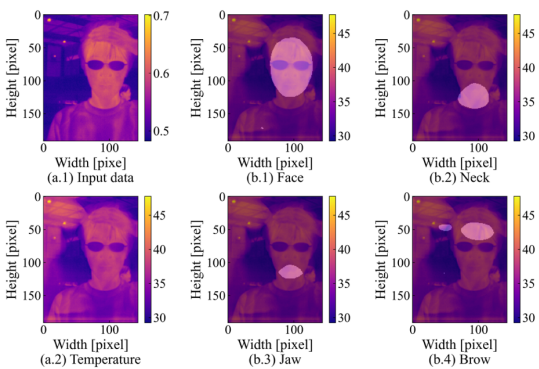


図-5 各学習モデルによる予測結果の一例

5. おわりに

本稿では、深層学習技術のひとつであるU-Netを用いてサーモグラフィデータの温度情報のみによる特定部位の検出と温度場の学習モデルの構築手法について検討した。

本稿において、学習モデルの検証時に過学習が確認されたが、この点に関してはデータ拡張により改善できると考えている。また、改善後の予測結果に対しては、IoU (Intersection over Union)による評価の実施を予定している。

謝辞: 本研究の一部は「愛知工科大学特別研究推進プロジェクト」および「摂南大学学生部研究助成金」の助成を受けて実施された。

参考文献

[1] Rowley, H. A., Baluja, S. and Kanade, T.: Human Face Detection in Visual Scenes, Advances in Neural Information Processing Systems 8, pp.875-881, 1995.

[2] Ukai, D. and Murakami, K.: An identification method using thermo-image, The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.155-158, 2013.

[3] Ronneberger, O., Fischer, P. and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention 2015, pp.234-241, 2015.