Anomaly detection in facial skin temperature using variational autoencoder

Ayaka Masaki¹ · Kent Nagumo¹ · Bikash Lamsal² · Kosuke Oiwa¹ · Akio Nozawa¹

Received: 15 April 2020 / Accepted: 16 August 2020 / Published online: 9 September 2020
© The Author(s) 2020

Abstract

Facial skin temperature is a physiological index that varies with skin blood flow controlled by autonomic nervous system activity. The facial skin temperature can be remotely measured using infrared thermography, and it has recently attracted attention as a remote biomarker. For example, studies have been reported to estimate human emotions, drowsiness, and mental stress on facial skin temperature. However, it is impossible to make a machine that can discriminate all infinite physiological and psychological states. Considering the practicality of skin temperature, a machine that can determine the normal state of facial skin temperature may be sufficient. In this study, we propose a completely new approach to incorporate the concept of anomaly detection into the analysis of physiological and psychological states by facial skin temperature. In this paper, the method for separating normal and anomaly facial thermal images using an anomaly detection model was investigated to evaluate the applicability of variational autoencoder (VAE) to facial thermal images.

Keywords Facial skin temperature · Thermal face image · Infrared thermography · Deep learning · Anomaly detection · Variational autoencoder

1 Introduction

Facial skin temperature is a physiological index that varies with skin blood flow controlled by autonomic nervous system activity [1]. The facial skin temperature can be remotely measured using infrared thermography, and it has recently attracted attention as a remote biomarker [2–4].

It has been known so far that there is some relationship between human physiological and psychological states and skin temperature at anatomical sites, but the appropriate size and position of region of interest (ROI) has not yet been clarified. Moreover, it is necessary to set an appropriate baseline according to the purpose of the experiment when

we evaluate the physiological and psychological state based on the facial skin temperature. Furthermore, there are many issues facing the practical application of facial skin temperature, such as the slow response of skin temperature due to changes in subcutaneous blood flow under existing conditions [1]. There are many studies that try to solve such the issues by methods such as pattern identification. For example, studies have been reported to estimate human emotions [5–8], drowsiness [9–11], mental stress [12], etc. In contrast to the preceding argument, it is impossible to make a machine that can discriminate all infinite physiological and psychological states. Considering the practicality of skin temperature, a machine that can determine the normal state of facial skin temperature may be sufficient. For example, the driver's drowsiness estimation is enough to determine that he is not sleepy. We only need to know that we are not stressed in case of stress judgment. Furthermore, we only need to know that we are healthy in case of a simple health check.

In this study, we propose a completely new approach to incorporate the concept of anomaly detection [13] into the analysis of physiological and psychological states by facial skin temperature. The proposed anomaly detection algorithm separates the normal facial skin temperature

This work was presented in part at the 25th International Symposium on Artificial Life and Robotics (Beppu, Oita, January 22–24, 2020).

✉ Ayaka Masaki
ojyarumarun@gmail.com
Akio Nozawa
akio@ee.aoyama.ac.jp

¹ Aoyama Gakuin University, Sagami-hara, Japan

² Kajima Corporation, Chofu, Japan

from the anomaly facial skin temperature such as “sleepy”, “stressed”, or “unhealthy”.

In the anomaly detection field, only normal data that can be collected easily are often used, since it is difficult to cover the data in the anomaly state. Therefore, we focus on the anomaly detection problem using unsupervised learning [14]. In the past, it was considered difficult to perform unsupervised learning in image space because of the curse of dimension, but the deep generative models [15] can deal with this problem. Typical deep generative models are autoencoder (AE) [16], variational autoencoder (VAE) [17], and generative adversarial network (GAN) [18]. In addition, many of these derivative technologies, vector quantized variational autoencoder-2 (VQVAE-2) [19, 20], anomaly detection with GANs (ADGAN) [21], and efficient GAN [22] have been reported, and anomaly detection in image space has made remarkable progress.

Among these algorithms, we investigate the VAE-based anomaly detection method for the following reasons. (1) The features of unlearned part are not reproduced. In other words, we cannot guarantee how it will behave for unlearned parts, since normal AE only learns how to reduce the learned dimension. (2) Convergence is guaranteed and learning is easier than GAN. (3) Latent variables can be dropped to a lower dimension, so it is easier to handle compared with GAN. (4) Easy to interpret because it has a mathematical structure [23]. VAE has also been applied to images in various wavelength bands such as detection of cancerous tissue parts from chest X-ray images [24] and anomaly detection of industrial products based on visible images [25]. Therefore, it is necessary to study whether VAE can be applied to infrared images that we specialize in.

In this study, we propose the anomaly detection model for facial skin temperature using VAE for the development of a machine that can judge the normal state of facial skin temperature. The objective of study is to detect anomaly facial thermal images from multiple facial thermal images. In this paper, the method for separating normal and anomaly facial thermal images using an anomaly detection model was investigated to evaluate the applicability of VAE to facial thermal images.

2 Anomaly detection using VAE

2.1 Overview of the VAE algorithm

VAE is a generative model based on deep learning. Figure 1 shows overview of the VAE algorithm. X , $\tilde{X}$, NN represent input, output, and a neural network, respectively. The VAE network is divided into an encoder section and a decoder

Fig. 1 Overview of VAE

section. Given observation $X = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N\}$, VAE identifies probability distributions ($p(\bar{x}_*|X)$) that produce unobserved value ($\bar{x}_*$). VAE is designed on the assumption that latent variables that serve as explanatory variables are normally distributed. The encoder performs dimensional compression of X , and it calculates the mean vector ($\bar{\mu}_\phi(x)$) and variance ($\bar{\Sigma}_\phi(x)$), which are parameters of the normally distribution. The blue part in Fig. 1 indicates that sampling is performed from the standard normal distribution. The points are then sampled from the latent space distribution ($\bar{z}$). In the decoder, the model likelihood parameters ($\bar{\eta}_\theta(z)$) are calculated, and the reconstruction error can be computed. Finally, the reconstruction error is backpropagated through the network. Since the reconstruction error to be optimized includes a regularization term that brings the mean ($\bar{\mu}_\phi(x)$) to 0 and the variance ($\bar{\Sigma}_\phi(x)$) close to the unit matrix, the distribution of the latent variable $\bar{z}$ has a shape close to a standard normal distribution. There is a tendency to regularize the organization of the latent space by bringing the distribution returned by the encoder closer to the standard normal distribution. For this reason, VAE can avoid overfitting and achieve a high recall compared to a normal autoencoder.

2.2 Anomaly detection using VAE

We explain the concept of anomaly detection in thermal face images using VAE. Figure 2 shows the conceptual diagram. Normally, with machine learning, it is necessary to have a dataset with a complete number of samples for each class. However, it is difficult to collect data when a person is abnormal (e.g., data on people who are not feeling well). Therefore, in this study, an evaluation of the model constructed using only the data in the normal condition, which is relatively easy to collect, was performed. In this study, we defined thermal face images in normal state as Normal, in anomaly state as Anomaly. Only Normal is used for anomaly detection using VAE. First, by learning a large amount of Normal using VAE, the anomaly detection model was constructed. Second, testing data (Anomaly and Normal) were input to the anomaly detection model. If the pattern was similar to Normal, the testing data were decided Normal; otherwise, the testing data were decided Anomaly.

Fig. 2 Conceptual diagram of anomaly detection using VAE

3 Proposed algorithm

3.1 Collection of anomaly and normal

Anomaly detection using VAE requires a large amount of normal data. In this study, we verify whether the proposed algorithm is effective in anomaly detection in facial skin temperature. Anomaly is thermal face images in which the skin temperature is forcibly changed by holding the breath. The Normal were thermal face images obtained when the subject was in the normal state. Image sizes are 640×480 pixels. The number of normal data was 4976, of which 90% was used as learning data and 10% was used as testing data. The number of Anomaly serving as the testing data was 195 in accordance with the testing data of Normal. The difference in the facial skin temperature value in a single image is small. The thermal images may also be biased by room temperature and other disturbances. In this study, the thermal face image was normalized, such that the maximum temperature was 1 and the minimum temperature was 0 for learning the relative amount of skin temperature inside the face. Specifically, the 640×480 pixel thermal image output from the thermographic device was cropped to 180×180

Fig. 3 Samples of learning data

Fig. 4 Overall of the anomaly detection model

Table 1 Construction of encoder

Layers	Function	Filters	Size	Stride
Conv1	—	16	2×2	2
BatchNorm1	ReLU	—	—	—
Conv2	—	32	2×2	2
BatchNorm2	ReLU	—	—	—
FC	—	—	—	—

pixels, leaving the face area. This thermal image after cropping is defined as a thermal face image. The normalization was applied independently to each of thermal face images.

3.2 Construction of the anomaly detection model

When performing VAE learning, a part of the thermal face image is randomly cut out at 8×8 size pixels, and this patch is used as learning data, as shown in Fig. 3. The local thermal image of the skin area was expanded to 10,000 sheets. Figure 4 shows overall of the anomaly detection model. In this study, convolutional layers were placed before the FC layer of encoder to extract the features of the skin temperature pattern of the skin blood vessels. Along with that, deconvolutional layers were placed after the decoder. The construction of the encoder is depicted in Table 1. The structure of the VAE encoder consisted of two convolutional layers and one fully connected layer. In this table, Conv, BatchNorm, and FC indicate the convolutional, batch normalization, and fully connected layers, respectively. The mean vector and

variance were output from FC. The structure of the decoder is paired with the structure of the encoder, and the structure is opposite to that of the encoder. The construction of the decoder is depicted in Table 2. ConvTrans indicates a transpose convolution. The gradient descent method was used for VAE parameter learning, and the optimization algorithm at that time was Adam. The number of dimensions of the latent variable $\bar{z}$ was searched from 4, 5, 6, 7, 8, 10, and the best model was selected. The number of epochs was 20 and the batch size was 128.

When testing, for all test data, the spatial unregularized anomaly score was calculated with reference to [25] and used as an index for detecting facial thermal images with some kind of abnormality from multiple facial thermal images. Unregularized anomaly score represented the error between the input vector and the reconstructed vector. To use the unregularized anomaly score to determine Anomaly, it is necessary to set a threshold mostly. However, the threshold is not defined in this study. In this paper, the statistics of Normal and Anomaly of unregularized anomaly scores were calculated to see whether it is possible to separate Normal and Anomaly. The equation for the spatial unregularized anomaly score is shown below:

$$L_{\text{VAE}}(x) = \sum_{i=1}^{N_x} \frac{1}{2} \times \frac{(\mu_{x_i} - x_i)^2}{\sigma^2 x_i} \bigg|_{z=\mu_z}. \quad (1)$$

The above equation is directly related to the reproduction error. N_x and x_i represent the number of pixels in the image and any pixel value, respectively. μ_{x_i} represents the maximum posterior probability estimation of the latent variable $\bar{z}$.

Table 2 Construction of decoder

Layers	Function	Filters	Size	Stride
FC	—	—	—	—
BatchNorm1	ReLU	—	—	—
ConvTrans1	—	32	2×2	2
BatchNorm2	ReLU	—	—	—
ConvTrans2	—	16	2×2	2
BatchNorm3	ReLU	—	—	—
ConvTrans3	—	1	4×4	1
ConvTrans4	—	1	4×4	1
BatchNorm4	ReLU	—	—	—
BatchNorm5	ReLU	—	—	—

4 Experiment

4.1 Experiment system

The experimental systems consisted of an infrared thermography device (FLIR A600-Series, FLIR systems Co., Ltd). The size of the thermal image was 640×480 pixels, and the temperature resolution was less than 0.05°C . The infrared emissivity of skin was $\epsilon = 0.98$. The viewing angle of the infrared thermography is 60° .

4.2 Procedure and condition

Healthy young subject (male; aged, 24) participated in the experiments. The subject provided informed consent about the experiments and objects of this study prior to agreeing to participate in the experiment. For the introduction to the real environment, subject was asked to cooperate in the experiment as usual as possible without modifying their daily activities such as food intake, sleep, smoking habits, etc. The experiment was conducted in an experimental room ($24.7 \pm 6.0^\circ\text{C}$). The ultimate goal of this study is to construct a system that recognizes the normal state of facial skin temperature for estimating driver drowsiness and checking stress in daily life. Therefore, we conducted the experiment in the room, which the environment temperature was not controlled, to simulate the daily life environment. In the experiment, the skin temperature of the face was measured twice. The sampling frequency of thermal images was 10 Hz. An infrared thermograph was placed 1 m in front of the subject. In the first time, the subject did not control anything, and they sat in a chair for about 10 min, with their face as still as possible. This thermal face image measured for about 10 min was regarded as Normal. The second time, the subject held his breath held for 50 s, which raised the blood pressure to promote fluctuations in facial skin temperature. The measurement was performed in iterations of 30–50 s, and the thermal face image obtained at this time was regarded as Anomaly.

5 Result and discussion

5.1 Anomaly and normal

Figure 5 shows examples of normalized thermal face image samples.

5.2 Anomaly detection

Figures 6 and 7 show the unregularized anomaly scores of two arbitrary samples. The unregularized anomaly scores are

Fig. 5 Samples of thermal face images

Fig. 6 Unregularized anomaly score of sample1

Fig. 7 Unregularized anomaly score of sample2

mapped on a log scale. The blue color indicates the degree of abnormality. When the unregularized anomaly scores were observed for all test data, they were categorized into two types of map, Figs. 6 and 7. In Fig. 6, the input thermal images are samples that vary significantly across facial skin temperature. When observing unregularized anomaly score, there is no abnormal part in the entire face in the normal and

Fig. 8 Histogram of unregularized anomaly score in sample1

Fig. 9 Histogram of unregularized anomaly score in sample2

abnormal parts in the nose, the left and right cheeks, and around the mouth in the anomaly space.

This is suggested that VAE learned a large number of samples in which no part of the entire face had a significant change in temperature. In Fig. 7, the nose was abnormal even in the normal space, but parts other than the nose were not recognized as abnormal. As a result, the VAE recognized a portion where the skin temperature was lowered as an abnormal portion and a portion where the skin temperature did not change as a normal portion. That is, the VAE algorithm was effective in capturing changes in temperature within the face. When constructing the anomaly detection model, we only use the thermal face images of Normal. In other words, the unregularized anomaly score calculated with the model that only learns Normal may be able to identify the state of the test image (Normal or Anomaly) even if it is not known whether the test image is Normal or Anomaly.

The histograms of unregularized anomaly score shown in Figs. 6 and 7 are shown in Figs. 8 and 9, respectively. The horizontal axis represents unregularized anomaly score and the vertical axis represents frequency of unregularized anomaly score. The vertical axis of these histograms is

Fig. 10 Average and variance of unregularized anomaly score

plotted on a log scale. For both samples, the general shape of the histogram distribution varied significantly. Abnormality patterns are categorized into two types, both of which have different distributions for anomaly and normal; therefore, the same can be said for all test data. Figure 10 shows the average and variance of the unregularized anomaly score for all test data. Since the anomaly space deviates from the normal space, the proposed algorithm is useful for detecting abnormal skin temperature fluctuations.

6 Conclusion

The objective of this study is to detect anomaly facial thermal images from multiple facial thermal images for the development of a machine that can judge the normal state of facial skin temperature. In this study, we proposed the anomaly detection model in thermal face images using VAE. In this paper, the method for separating normal and anomaly facial thermal images using an anomaly detection model was investigated to evaluate the applicability of VAE to facial thermal images. In actual fact, we collected the Normal skin temperature and the anomaly skin temperature by experiment. Using this anomaly detection model to calculate unregularized anomaly score, different distributions of unregularized anomaly score were obtained between Normal and Anomaly in all test data. As a result, it has been shown that VAE can be used to detect abnormalities in facial skin temperature. In the future, we plan to use probability models to discriminate between Normal and Anomaly, and perform quantitative evaluations using discrimination rates. In addition, we plan to increase the number of samples and conduct an assessment of generality.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Loannou S, Gallese V, Merla A (2014) Thermal infrared imaging in psychophysiology: potentialities and limits. *Psychophysiology* 51(10):951–963
- Kan H, Liu G (2017) Facial thermal image analysis for stress detection. *Int J Eng Res Technol* 6(10):94–98
- Nakane N, Oiwa K, Nozawa A (2020) Relationship between mechanisms of blood pressure change and facial skin temperature distribution. *Artif Life Robot* 25(1):48–58
- Oiwa K, Okamoto R, Bando S, Nozawa A (2018) 'Blind source extraction of long-term physiological signals from facial thermal images. *Artif Life Robot* 23(2):218–224
- Zenju H, Nozawa A, Ide H (2004) Estimation of unpleasant and pleasant states by nasal thermogram. *IEEE J Trans Electron Inf Syst* 124:213–214 (in Japanese)
- Hisaya T, Ide H, Nagashuma Y (2000) An attempt of feeling analysis by the nasal temperature change model". *Smc 2000 conference proceedings*. In: 2000 IEEE international conference on systems, man and cybernetics. 'cybernetics evolving to systems, humans, organizations, and their complex interactions', cat. no. 0, vol 2 IEEE, pp 1265–1270
- Nakanishi R, Imai-Matsumura K (2008) Facial skin temperature decrease infants with joyful expression. *Infants Behav Dev* 31(1):137–144
- Sjoerd J, Ebisch A, Aureli T, Bafunno D, Cardone D, Romani GL, Merla A (2008) Mother and child in synchrony: thermal facial imprints of autonomic contagion. *Biol Psychol* 89(1):123–129
- Hirotsoshi A, Naoki S, Nozawa A, Ide H (2010) Presumption of transient awakening of driver by facial skin temperature. *IEEE J Trans Electron Inform Syst* 130(3):428–432 (in Japanese)
- Adachi H, Oiwa K, Nozawa A (2019) Drowsiness level modeling based on facial skin temperature distribution using a convolutional neural network. *IEEE J Trans Electric Electron Eng (TEEE C)* 14(6):870–876
- Bando S, Oiwa K, Nozawa A (2017) Evaluation of dynamics of forehead skin temperature under induced drowsiness. *IEEE J Trans Electric Electron Eng* 12(S1):S104–S109
- Veronika E, Arcangelo M, Grant JA, Daniela C, Tusche A, Singer T (2014) Exploring the use of thermal infrared imaging in human stress research. *PLoS One* 9(3):125–136
- Varun C, Banerjee A, Kumar V (2009) Anomaly detection: a survey. *ACM Comput Surv (CSUR)* 41(3):1–58
- Hofmann T (2001) Unsupervised learning by probabilistic latent semantic analysis. *Mach Learn* 42(1–2):177–196
- Rezende DJ, Mohamed S, Wierstra D (2014) Stochastic back-propagation and approximate inference in deep generative models. [arXiv:1401.4082](https://arxiv.org/abs/1401.4082)

16. Sakurada M, Takehisa Y (2014) 'Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis, pp 4–11
17. An J, Cho S (2015) Variational autoencoder based anomaly detection using reconstruction probability, special lecture on IE 2.1
18. Schlegl T, Seebock P, Waldstein SM, Schmidt-Erfurth U, Langs G (2017) Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International conference on information processing in medical imaging IPMI2017, pp 146–157
19. Razavi A, van den Oord A, Vinyals O (2019) Generating diverse high-fidelity images with vq-vae-2. *Adv Neural Inf Process Syst* 14:14837–14847
20. Zimmerer D, Petersen J, Maier-Hein K (2019) High-and Low-level image component decomposition using VAEs for improved reconstruction and anomaly detection. [arXiv:1911.12161](https://arxiv.org/abs/1911.12161)
21. Deecke L, Vandermeulen R, Ruff L, Mandt S, Kloft M (2018) Anomaly detection with generative adversarial networks. [arXiv:1809.04758](https://arxiv.org/abs/1809.04758)
22. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR (2018) Efficient gan-based anomaly detection. [arXiv:1802.06222](https://arxiv.org/abs/1802.06222)
23. Lu Y, Xu P (2018) Anomaly detection for skin disease images using variational autoencoder. [arXiv:1807.01349](https://arxiv.org/abs/1807.01349)
24. Kurotaki H, Nakayama K, Uehara M, Yamaguchi R, Kawazoe Y, Ohe K, Matsuo Y (2017) Diagnosis support from chest X-ray pictures with deep network. In: The 31st annual conference of the Japanese society for artificial intelligence, 2017, 2B1-3 (**in Japanese**)
25. Tachibana R, Matsubara T, Uehara K (2018) Anomaly manufacturing product detection using unregularized anomaly score on deep generative models. In: The 32nd annual conference of the Japanese society for artificial intelligence, 2018, 2A1-03 (**in Japanese**)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.