Electronic Preprint for Journal of Information Processing Vol.26

Regular Paper

Data Augmentation to Build High Performance DNN for In-bed Posture Classification

Yu Enokibori[1,a)] Kenji Mase[1,b)]

Received: December 22, 2017, Accepted: July 10, 2018

Abstract: E-textiles have come to be used instead of several types of common equipment, such as bed-sheets, in some cases. An application using body pressure data collected through such bed-sheet type sensors is the in-bed posture classification expected for pressure ulcer prevention. Since such body pressure data is a kind of low-resolution image, Deep Neural Network (DNN) based algorithms seem suitable. However, it is difficult to collect enough data to use for DNN in this domain because the number of sleep postures obtained from one experiment is small. For an example, the number of postures collected from 19 subjects with four hours of sleep each is only 224. To solve such a small data-size problem in DNN, data augmentation techniques have been proposed. However, random augmentations are not so suitable. Therefore, we investigated appropriate augmentation parameters for this domain. As a result, the combination of the up to ±20% and ±40% random shifts along short and long sides of a bed, the up to ±10 degree rotation, and non-use of other transformations showed the best performance. With the parameters, the built DNN showed 99.7% accuracy and 0.997 Weighted F^1-score for three posture classifications: supine, left and right lateral positions, and 97.1% accuracy and 0.970 Weighted F^1-score for four posture classifications: supine, prone, left and right lateral positions.

Keywords: data augmentation, deep neural network, in-bed posture classification, pressure ulcer prevention, e-textile, pressure sensor

1. Introduction

Performance of e-textiles has improved, and such e-textiles have come to be used instead of several types of common equipment, such as bed-sheets, in some cases. An example application using body pressure data during sleep collected through such bed-sheet type sensors is the in-bed posture classification expected for pressure ulcer prevention. An example of body pressure data and our bed-sheet-type pressure sensor is shown in **Fig. 1**.

Since such body pressure data is a kind of low-resolution image, Deep Neural Network (DNN) based algorithms [1], [2], [3] seem to be suitable for classifier implementation. In addition, high performance algorithms that seem useful for pressure ulcer prevention are developed on DNN actively, such as human body-part detection [4], [5], [6] and human posture detection with occlusion avoidance [7], [8]. These algorithms can be used for tracking of high risk parts of the body, in-bed posture estimations avoiding disturbance of pressure dispersion cushions, and so on.

However, it is difficult to collect enough data to use for DNN training of sleep posture classification because the number of sleep postures obtained in one experiment is small. Moreover, data collected from natural sleeping includes many variations in one posture, such as arm and leg positions, angles of elbow and knee joints, and overlapping body parts. Showing an overview of our experiment as an example, the number of postures collected from 19 subjects with four hours of natural sleep each is only 224 totally: 118 supine positions from 16 subjects, 51 left lateral positions from 13 subjects, 40 right lateral positions from 15 subjects, and 15 prone positions from six subjects. Typical examples of the supine, lateral and prone postures are shown in **Fig. 2**. In addition, these included many variations in one posture, such as inserting hands under a pillow, folding legs as in the shape of the number “4”, and so on. Several examples of such variations are shown in **Fig. 3**. These sample sizes of classes are too small for DNN training, compared with typical DNN studies such as

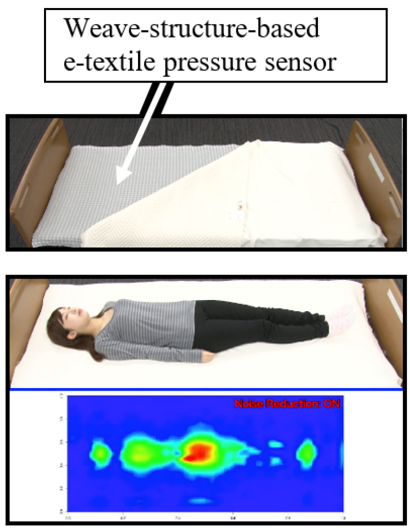

Fig. 1 Bed-sheet type e-textile pressure sensor and example data.

1 Nagoya University, Nagoya, Aichi 464–8603, Japan
a) enokibori@i.nagoya-u.ac.jp
b) mase@nagoya-u.jp

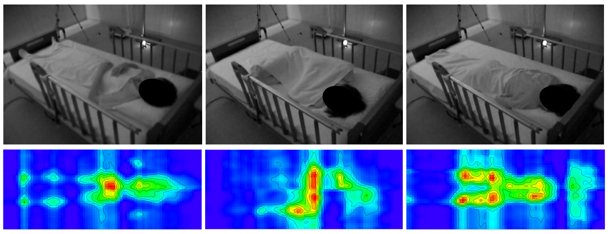

Fig. 2 Examples of typical supine, lateral and prone postures.

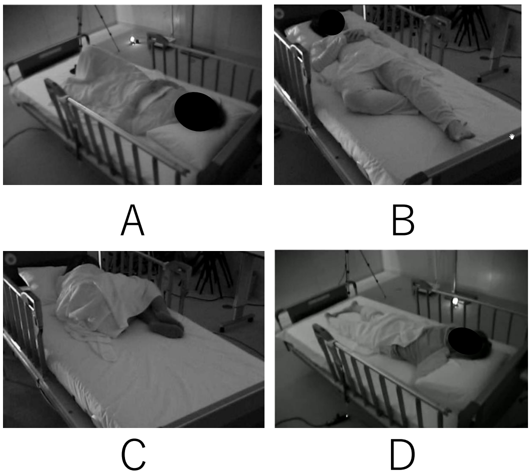

Fig. 3 Examples of posture variety (A: rising knees, B: "4" letter like legs, and the right foot hiding under the left leg, C: hard curl, D: arms inserted under the pillow).

Refs. [1], [2], [3].

To solve such a small data size problem in DNN, data augmentation techniques have been proposed [9], [10]. Such techniques augment data size with post processing, such as expanding, cropping and rotating. However, random augmentations are not so suitable. Investigations of appropriate augmentation parameters for each data domain are necessary and important.

Therefore, we are investigating appropriate augmentation parameters for DNN in this domain, to enhance the performance of DNN based algorithms that seem useful for pressure ulcer prevention, such as human body-parts detection and human posture detection with occlusion avoidance that are described already. This paper, as the first step to establish such DNN based pressure ulcer preventions, describes a result of investigation with a simple in-bed posture classifier that is also not studied enough.

We used only simple methods to data augmentation in this investigation because complex methods, such as multi-scale sliding window, may provide bad effects and conflicts for such related algorithms. The investigated parameters are ones of shear transformation, zoom, rotation, and shifts along short and long sides of a bed.

As a result, the combination of the up to ±20% and ±40% random shifts along short and long sides of a bed, the up to ±10 degree rotation, and non-use of other transformations showed the best performance. With the parameters, the built DNN showed 99.7% accuracy and 0.997 Weighted F^1-score for three posture classifications: supine, left and right lateral positions, and 97.1% accuracy and 0.970 Weighted F^1-score for four posture classifications: supine, prone, left and right lateral positions.

The rest of the paper is organized as follows. Section 2 summarizes related works. Section 3 presents our dataset overview. Section 4 describes the details of our data augmentation parameter investigation. Section 5 shows the performance of DNN trained with the investigated data augmentation parameters. Section 6 discusses the comparison result with related works. Finally, Section 7 concludes the paper.

2. Related Works

In-bed posture classifications using pressure sensors are studied well as in the following. However, most of them are based on SVM (Support Vector Machine) and kNN (k-Nearest Neighbor). DNN is not so significant in this domain.

In-bed posture classifications based on SVM were presented in the following studies. Hsia et al. established 83.5% accuracy for six posture classifications, not including the prone position, using a force-sensitive resistor (FSR) sensor array [11]. Huang et al. established 94.1% accuracy for nine posture classifications, including the prone position, with data collected from FSR sensors and cameras [12]. Mineharu et al. established 77.1% accuracy for nine posture classifications, including the prone position, with matrix pressure data and region depending on feature extraction [13]. Hayashi et al. established 92.3% accuracy for four posture classifications, including the prone position, and 96.3% accuracy for three posture classifications, not including the prone position, with matrix pressure data and of Center-of-Pressure features [14].

In-bed posture classifications based on kNN were presented in the following studies. Yousefi et al. established 97.7% accuracy for five posture classifications, not including the prone position, using PCA (principal component analysis), kNN and binarized matrix pressure data [15]. Pouyan et al. established 97.1% accuracy for eight posture classifications, not including the prone position, using kNN with humming distance and binarized matrix pressure data [16]. Ostadabbas et al. established 98.4% accuracy for three posture classifications, not including the prone position, using the combination of kNN and GMM [17].

A DNN based study was presented by Heydarzadeh et al. [18]. They used HoG based features for input of an auto-encoder type DNN, and established 98.1% accuracy for five posture classifications, not including the prone position. They also used data augmentation, but there is no detailed discussion about appropriate parameters of the data augmentation. Moreover, use of HoG based features would interfere with collaborations related with DNN techniques. More basic DNN, as shown in this paper, is required for such collaborations.

3. Dataset Overview

In this section, we describe overview of our dataset.

3.1 Sensor

The data was collected with an e-textile pressure sensor shown in **Fig. 4**. Its specification is summarized in **Table 1**. This sensor is a kind of weave-structure based sensor. Our textile sensor has multiple capacitance sensors between the weft and the warp in

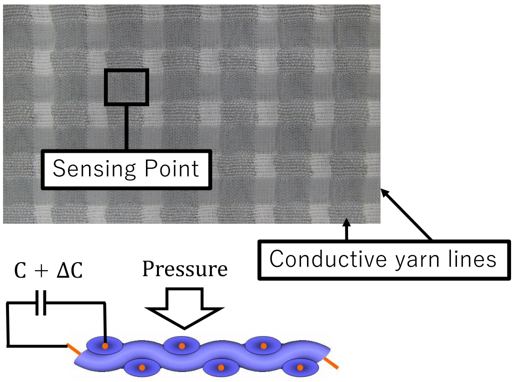

Fig. 4 Overview of textile-weave-structure based pressure-sensor.

Table 1 The specification of the e-Textile sensor.

Size	180×90 cm (depending on the setup)
Thickness	0.6 mm
Sensing point size	1 cm^2
Distance of sensing points	$2 \sim 2\sqrt{2}$ cm (between their center)
Resolution	3,200 (80×40)
Sampling rate*	up to 10 Hz per one surface
Interface	Wi-Fi / Bluetooth

* Sampling rate was set as 2 Hz in this study.

Table 2 Characteristics of subjects.

	elderly	young
Age	68.8 ± 4.0	21.3 ± 0.8
Male	6	6
Female	4	6
Body Mass Index (BMI)	21.8 ± 3.7	21.3 ± 3.0

* All subjects were healthy.

the cross points of the gray conductive yarn lines. Values of the capacitance sensors change depending on the distance between the weft and the warp that is altered if additional load is added to the textile surface. Thus, we can measure the size of the load that is placed on the textile through capacitance value change measurements using a capacitance measuring device. The sensor was formed as a bed-sheet for a single bed whose size was 180×90 cm in this study. Its sampling rate was set as 2 Hz per surface. The size of sensing points was 1 cm square. Such sensing points were located as a matrix pattern. The distance between centers of adjacent sensing points was 2 cm.

3.2 Subject

Subjects were 22 healthy elderly and young people as summarized in **Table 2**. The average age of the elderly group (6 males and 4 females) was 68.8 ± 4.0. The average age of the young group (6 males and 6 females) was 21.3 ± 0.8. Body mass indexes (BMI) of the elderly and young groups were 21.8 ± 3.7 and 21.3 ± 3.0. This means they had standard body shape. Unfortunately, the data on three young subjects (2 males and 1 female) included errors related to machine troubles. Therefore, this study excluded such subjects from the dataset, and used the data of 19 healthy elderly and young subjects.

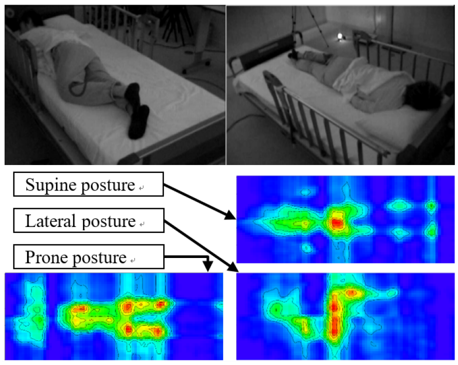

Fig. 5 Overview of body-pressure data collection.

3.3 Process Overview and Sample Data

Figure 5. is an overview of the data collection environment and samples of collected data. We installed the bed-sheet type pressure sensor in the same manner as Fig. 1 and put two infrared cameras to collect ground truth data. The lights of the room were turned off. The temperature of the room was controlled to be comfortable. Subjects slept four hours each naturally.

The heat-maps in the lower side of Fig. 5 are visually determinable examples of collected data. The dataset also includes many visually indeterminable ones because they are collected from natural sleeping. Moreover, the dataset includes many variations in one posture, such as supine and lateral positions with their hands inserted under a pillow, supine positions with their legs folded like letter “4”, supine positions with raised knees, lateral positions with a hand protruding from the bed, and so on.

3.4 Posture Data Extraction and Post Processing

The posture data was extracted in the following manner. First, we extracted stable postures with no body movement of more than five minutes. Then, the first frame one minute after the body movement stopped was extracted as the data of the posture. This study used only one data for one posture because there are no significant sensor output changes during one stable posture. This means that the data of the same posture can not be included in training, validation, and test dataset at the same time even if naive data separations such as simple cross validation methods were used. The number of collected postures was 224 with 118 supine positions from 16 subjects, 51 left lateral positions from 13 subjects, 40 right lateral positions from 15 subjects, and 15 prone positions from 6 subjects. The data was doubled with flipping along the short side of the bed. Thus, finally, the number of collected postures was 448 with 236 supine positions from 16 subjects, 91 left lateral positions from 16 subjects, 91 right lateral positions from 16 subjects, and 30 prone positions from 6 subjects.

4. DNN Based Classifier for In-bed Postures

Figure 6 is the DNN used for in-bed posture classification in this study. Its input was arranged as 28×28 to be able to be applied for more sparse sensors. Thus, our sensor data was subsampled to 28×28 from 80×40 first. The input data was processed through the two 3×3 convolution layers whose kernel sizes were 32 and 64. Then, it was processed through a 2×2 max pooling

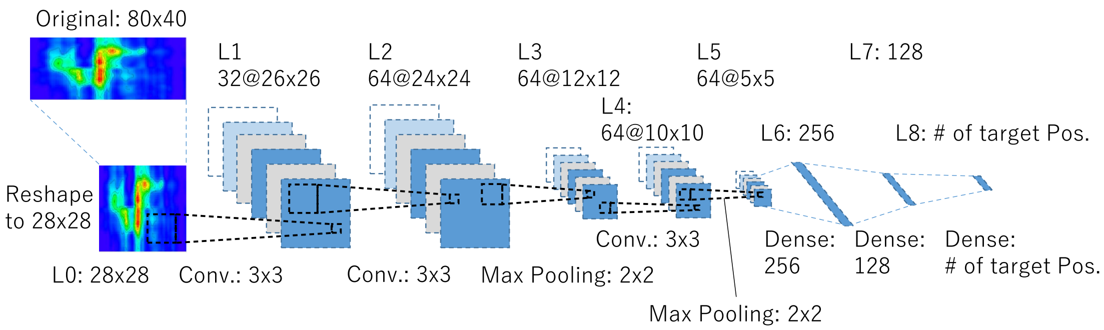

Fig. 6 CNN based DNN for in-bed posture classification.

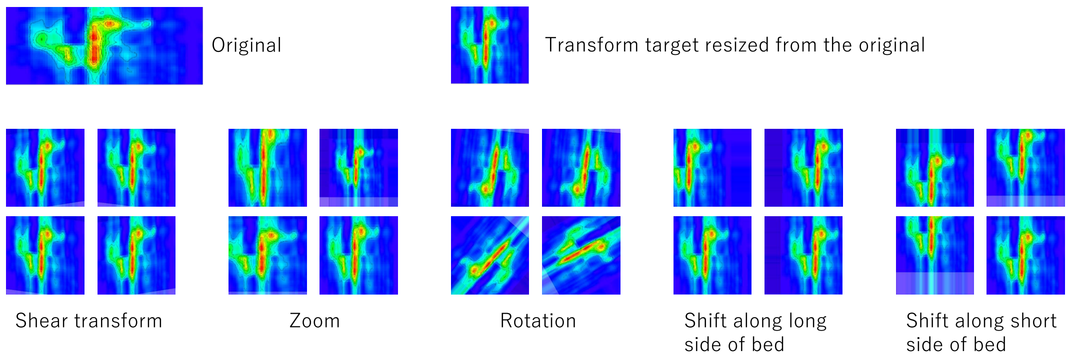

Fig. 7 Example images transformed by data augmentation.

layer for position aberration adjustment. It was once again processed through a 3×3 convolution layer whose kernel size was 64 and a 2 × 2 max pooling layer. Finally, it was calculated through the two dense layers whose kernel sizes were 256 and 128. The output layer was a dense layer with the kernel size of the number of target postures. The activation functions used were: Softmax for the output layer (L8) and LeRU for the others. The optimizer used was Nadam.

5. Appropriate Parameter Investigation for Data Augmentation

We explored the best combination of data augmentations for in-bed posture classification domain from the following transformations: shear transformation, zoom, rotation, and shifts along short and long sides of the bed. Examples of transformed images were shown in **Fig. 7**.

The shear and zoom transformations were expected to solve body shape and size variety insufficient. The rotation, and shifts along short and long sides of a bed transformations were expected to solve position variety insufficient. In addition, combinations among them are expected to solve miscellaneous issues. For example, combinations of rotation, zoom and shifts were expected to solve variety insufficient of free angle body parts, such as arms and legs. This is because the investigation range of the rotation have 0 to 360 degree of the arc despite of the fact that adults sleep within 5 or 10 degree rotations in most cases as the bed is narrow. Such small degree rotations are investigated additionally in Section 5.4.

We note that the above transformations are limited to the geometrical transformations. This means the data augmentation will not increase posture variety. At least, one data is required for each posture.

5.1 Investigation Ranges of Parameters

The investigation ranges of the parameters of the transformations were as follows.

- shear transformation: ±0, 0.2, 0.4, 0.6
- zoom: ±0, 0.2, 0.4, 0.6
- rotation: ±0, 90, 180, 270, 360 degrees of arc
- shift along long side of bed: ±0, 0.2, 0.4, 0.6 of long side length
- shift along short side of bed: ±0, 0.2, 0.4, 0.6 of short side length

The total number of combinations was 1,280. The transformations were applied with a random parameter between the selected limits. The individual transformations were applied for one input, multiply.

5.2 Investigation Process

We investigated the best parameter with the DNN trainings for the three posture classifications: supine, and right and left lateral postures. The subjects of even number IDs were excluded in this parameter investigation. A half of the subjects are used for training, and the others for validation. Both datasets included almost equal numbers of the elderly and young subjects. The data of the prone position was excluded in this investigation because the data was so small and disturbed healthy DNN training.

The data augmentations were applied only for the training dataset. The training parameters were the following.

- batch size: 64
- batch iterations per one epoch: 200
- total epoch: 30

We note that the meaning of the above epoch is not the same as common machine learning terminology that is a cycle to use up all the training dataset because "a cycle to use up all the training dataset" cannot be defined with the data augmentation described in this paper. The data augmentation provides new training data unlimitedly. Therefore, we defined the epoch as 64 batches × 200 iterations = 12,800 trainings with generated data, and applied common processes that should be run every epoch, such as learning rate adjustment.

Thus, a total of 384,000 transformed data were learned in one training. Source data for the data augmentation were selected randomly and removed from the source dataset. When all data are selected from the source dataset, the selection was restarted with the default source dataset. The epoch result with the lowest validation loss was selected as the best model of each training. Two trainings were done for each combination of investigated parameters, and each mean value of training and validation losses of the best model of each of the two trainings was used for the parameter combination selection described in the next section.

5.3 Investigation Result and Discussion

In this study we used mean values of training loss and validation loss for the best parameter selection, compared with typical DNN studies, because several training results showed low validation loss but also high training loss. Such combinations of low validation loss and high training loss mean such trainings were not processed correctly.

Figure 8 is the whole summary of the parameter investigation sorted in ascending order. The result on the left of the graph is better. **Figure 9** is a detailed summary of the top 10 results of

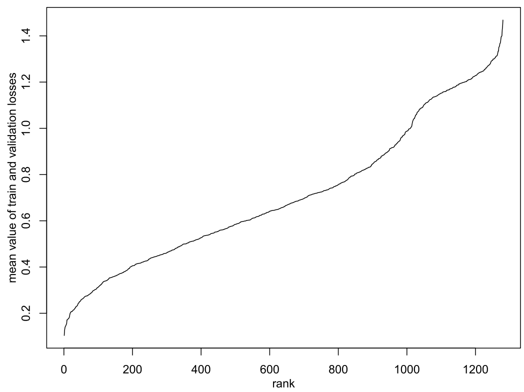

Fig. 8 Mean values of train and validation losses of all training.

Fig. 8. Each mark on each entry in Fig. 9 is the parameter of:

- shear transformation: red circle and dashed line
- zoom: green circle and dotted line
- shift along short side of bed: blue circle and chain line
- shift along long side of bed: light blue circle and long dashed line
- rotation: purple circle and long chain line
- mean value of training and validation losses: black circle and solid line

As shown in Fig. 8, selection of parameter combinations of data augmentation greatly affected the construction results of classifiers. As shown in Fig. 9, rotation was not used in all cases of top 10 results. This means such big degree rotations are not useful on the data augmentation in this domain. Zoom was not used in seven out of the top 10 results. This means zoom is not so useful on the data augmentation in this domain. On the other hand, the ±0.2 parameter for shift along the short side of the bed was used in all of the top five results and eight out of the top 10 results. This means the parameter value of the transformation greatly affects usefulness of the data augmentation in this domain. On the rest of the parameters, results of shift along the long side of the bed and shear transformation showed no significant trend. There are five ±0.2 and five ±0.4 parameters for shift along the long side of the bed, and there are all parameter variations are shown for the shear transformation. In this paper, we selected ±0.4 parameter for shift along the long side of the bed because the parameter appeared in the best result and also appeared in three out of the top five results. With shear transformation, we selected ±0.0 appeared in the best result as the best parameter in this paper. This means shear transformation should not be used.

As the above result, the most variety insufficient of data in this domain seems to be depending on the sleeping positions. Body shape and size are not a big matter or these are avoided with a dataset that has variety of sleeping positions. In addition, several combination effects expected did not appear in the result such as a combination of rotation, zoom and shifts to solve variety insufficient of free angle body parts, such as arms and legs.

5.4 Detailed Investigation for Rotation

In this section, we additionally investigate the best range of ro-

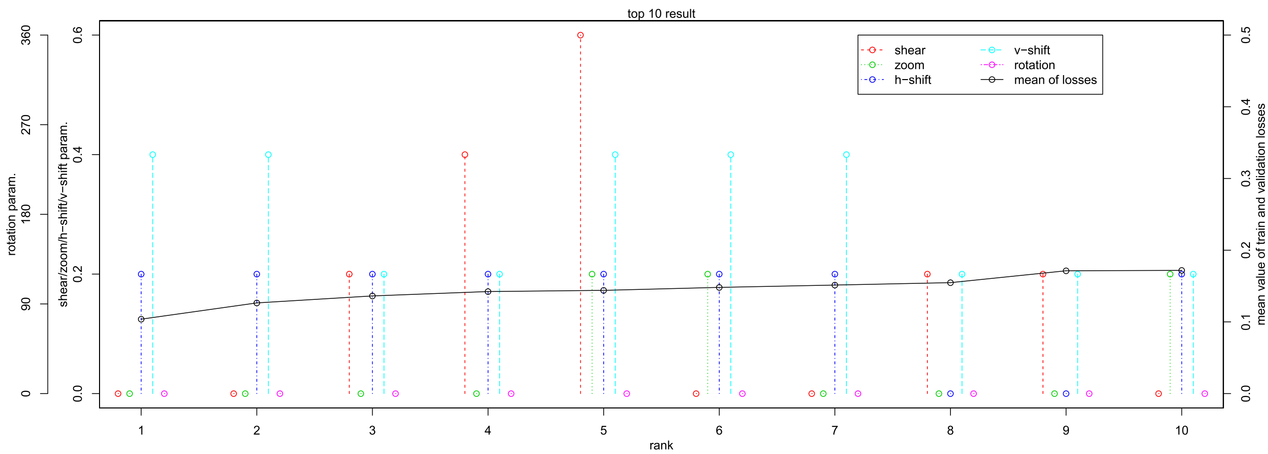

Fig. 9 Top 10 results of parameter investigation.

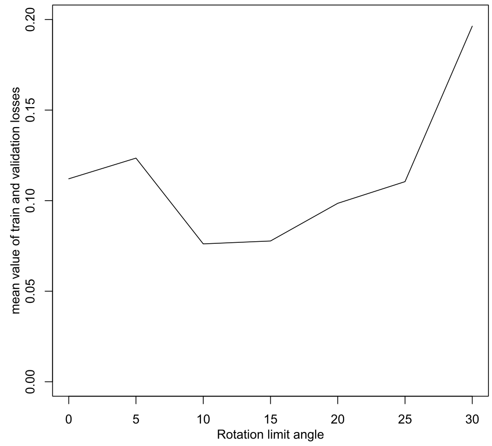

Fig. 10 Mean values of train and validation losses of small degree rotations.

tation transformation with small degrees, such as 0, 5, 10, 15, 20, 25 and 30. Adults sleep within 5 or 10 degree rotations in most cases as the bed is narrow. Therefore, the best parameter of rotation transformation is expected to be around such degrees instead of big degrees such as 360, if combinations of rotation and other parameters do not provide good effects.

In this investigation, parameters except rotation were fixed for the best parameters obtained in the previous section. This means the combination of no shear, no zoom, ±0.4 for shift along the long side of the bed, and ±0.2 for shift along the short side of the bed.

Figure 10 is an investigation result. As it was expected, the 10 degree showed the best performance. Therefore, we selected the 10 degree as the best parameter of rotation transformation.

6. Performance of DNN with the Best Data Augmentation Parameters

As a result of Section 5, the combination of the up to ±20% and ±40% random shifts along the short and long sides of the bed, the up to ±10 degree random rotation, and the non-use of other transformations showed the best performance. Therefore, we built DNN with the parameter combination and evaluated its performance.

The evaluations were done in the following cross validation (CV) manner.

- Fixed data for training: odd number id subjects
- CV data for training, validation and test: 1/3 of even number id subjects (d_0, d_1, d_2)

Therefore, in total six validations were in one CV. We tried ten trainings for each validation of CV and selected a training result that has the lowest validation loss as the best model of the validation. This is because the proposed approach, data augmentation with small base dataset, have instability. Many combinations of selected randomized parameters and base data do not provide significant improvement for trainings. If such non-effect combinations get a majority of training data, the training result show low performance even if several good combinations are provided in the last part of the training because the training process used several common techniques, such as learning rate reductions with continuous non-improvement trainings. Therefore, we select the best result from 10 trainings to solve such instability.

Table 3 Accuracy of each training w/o D.A for 3 Pos.

Data			
Training	Validation	Test	Accuracy
d_0	d_1	d_2	0.759
d_1	d_2	d_0	0.966
d_2	d_1	d_0	0.966
d_0	d_2	d_1	0.976
d_1	d_0	d_2	0.778
d_2	d_0	d_1	0.976
	mean ± SD		0.903 ± 0.095

Table 4 Confusion matrix w/o D.A for 3 Pos.

		Estimation				
		LL	RL	SP	recall	F^1
True	LL	67	3	6	0.882	0.865
	RL	3	69	4	0.908	0.885
	SP	9	8	219	0.928	0.942
precision		0.848	0.863	0.956		
				CV Accuracy	0.916	
				Weighted F^1-score	0.915	

LL: Left Lateral, RR: Right Lateral, SP: Supine

The test results were calculated with the best models, such as TP (true positive), FP (false positive), TN (true negative), FN (false negative), precision (TP/(TP + FP)), recall (TP/(TP + FN)), accuracy (TP/(TP + FP + TN + FN)) and F^1-score ($2 \cdot$ precison $\cdot$ recall/(precison + recall)). We also used Weighted F^1-score to fix the inbalance of the sample size of each class. Its definition is the following.

$$\text{Weighted F}^1\text{-score } (F^1_t) = \sum_i w_i \mathrm{F}^1_i$$

$$w_i = n_i/N$$

where i: i-th class,
F^1_i: F^1-score of i-th class,
n_i: sample size of i-th class,
N: total size of samples.

The DNN in this study was, first, trained with only three postures: supine, and right and left lateral positions. Then, it was trained for four postures with a fine-tuning approach. This is because the amount of prone position data was small. In the rest of this section, we will discuss these separately.

6.1 Comparison with Three Posture Classifications

The results without data augmentation are summarized in **Table 3** and **Table 4**. The results with data augmentation with the best parameters are summarized in **Table 5** and **Table 6**.

The significant accuracy improvements are shown in comparison between Table 3 and Table 5. There is the same trend in the comparison between Table 4 and Table 6. The values of CV accuracy and Weighted F^1-score in each confusion matrix table are almost the same. This means the trained classifiers were not trained biased to class with a large sample size.

Since the comparison between Table 3 and Table 5, the biggest improvement occurred when test data is d_2. **Figure 11** shows several examples that were estimated as wrong classes without data augmentation and were estimated as correct classes with data augmentation. The left one is shifted for the upper side a lot, thus the parameter of shifts make such improvement. In the

Table 5 Accuracy of each training w/ D.A for 3 Pos.

Data			
Training	Validation	Test	Accuracy
d_0	d_1	d_2	1.000
d_1	d_2	d_0	1.000
d_2	d_1	d_0	1.000
d_0	d_2	d_1	1.000
d_1	d_0	d_2	1.000
d_2	d_0	d_1	0.988
mean ± SD			0.998 ± 0.005

Table 6 Confusion matrix w/ D.A for 3 Pos.

		Estimation				
		LL	RL	SP	recall	F^1
True	LL	76	0	0	1.000	1.000
	RL	0	75	1	0.987	0.993
	SP	0	0	236	1.000	0.998
precision		1.000	1.000	0.996		
				CV Accuracy	0.997	
				Weighted F^1-score	0.997	

LL: Left Lateral, RR: Right Lateral, SP: Supine

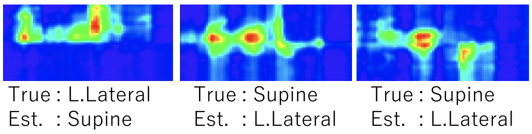

Fig. 11 Examples of wrong estimation w/o D.A. when test data is d_2.

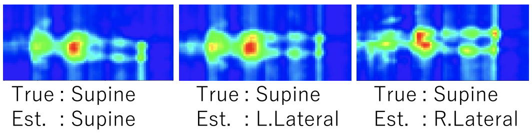

Fig. 12 Similar data of correct and wrong estimations w/o D.A. when test data is d_2.

middle one, subject's legs folded as letter "4". This is because it was miss-categorized as left lateral posture. On the other hand, why it is categorized into correct class with data augmentation is not clear. The improvement seems to be caused by only the enhancement of the training data number because shifts and rotation transformation do not seem to make effect. The right side example seems to be improved with rotation transformation. Unfortunately, these three example are the most meaningful data that human can understand. Most causes of improvements cannot be specified. **Figure 12** were other examples. The left one was estimated for correct class, however, the rest that are very similar for the left one were categorized into wrong classes. The enhancement of the training data size seems to be the main cause of such improvement, however, it is not clear. It should be discussed in future work.

6.2 Comparison with Four Posture Classification

The results without data augmentation are summarized in **Table 7** and **Table 8**. The results with data augmentation with the best parameters are summarized in **Table 9** and **Table 10**. Note here again, the classifiers for four postures were trained with a fine-tuning approach using the best model of each validation of three posture classification CV. The output layer was replaced and re-trained with the whole dataset. This is because the number of prone position data was small.

Table 7 Accuracy of each training w/o D.A. for 4 Pos.

Data			
Training	Validation	Test	Accuracy
d_0	d_1	d_2	0.645
d_1	d_2	d_0	0.950
d_2	d_1	d_0	0.933
d_0	d_2	d_1	0.859
d_1	d_0	d_2	0.613
d_2	d_0	d_1	0.859
mean ± SD			0.810 ± 0.133

Table 8 Confusion matrix w/o D.A for 4 Pos.

		Estimation					
		LL	RL	SP	PR	recall	F^1
True	LL	65	3	8	0	0.855	0.798
	RL	3	67	6	0	0.882	0.793
	SP	7	12	217	0	0.919	0.897
	PR	8	8	14	0	0	na
precision		0.783	0.744	0.886	na		
					CV Accuracy	0.835	
					Weighted F^1-score	na	

LL: Left Lateral, RR: Right Lateral, SP: Supine, PR: Prone

Table 9 Accuracy of each training w/ D.A. for 4 Pos.

Data			
Training	Validation	Test	Accuracy
d_0	d_1	d_2	0.968
d_1	d_2	d_0	0.983
d_2	d_1	d_0	0.981
d_0	d_2	d_1	0.978
d_1	d_0	d_2	0.952
d_2	d_0	d_1	0.978
mean ± SD			0.973 ± 0.012

Table 10 Confusion matrix w/ D.A for 4 Pos.

		Estimation					
		LL	RL	SP	PR	recall	F^1
True	LL	75	0	1	0	0.987	0.962
	RL	0	76	0	0	1.000	0.993
	SP	1	0	232	3	0.983	0.985
	PR	4	1	2	23	0.767	0.821
precision		0.938	0.987	0.987	0.885		
					CV Accuracy		0.971
					Weighted F^1-score		0.970

LL: Left Lateral, RR: Right Lateral, SP: Supine, PR: Prone

Here are the same trends with the results of three-posture classification. The significant accuracy improvements are shown in comparison between Table 7 and Table 9, and the comparison between Table 8 and Table 10. The values of CV accuracy and Weighted F^1-score in each confusion matrix tables are almost the same. This means the trained classifiers were not trained biased to class with a large sample size. Although the CV accuracy shown in Table 10 is lower than ones shown in Table 6, it still shows high performance value of 97.1%.

Except the improvement related to d_2 test data, since the comparison between Table 8 and Table 10, the biggest improvement occurred for prone postures. **Figure 13** shows several examples of prone data with estimation results using data augmentation. The left and middle ones were examples classified into correct class with the trainings with data augmentation. The right one is an example classified into incorrect class with the trainings with data augmentation. The left and middle ones have similar shape and position difference. 24 out of 30 prone data have such charac-

Table 11 Accuracy comparison with related works.

Cite	Sensor	# of sensing points	Classifier	Acc.	# of Pos.	Including prone pos.	# of Sbj.	# of data	Pos. Instruction	Variety in positions	Clear data exclusion of the same posture among training, validation and test data
[11]	Pressure (FSR)	56	RawData + SVM	83.5	6	NO	8	43,200	YES	NO	YES (3 sbj. for training / 5 sbj. for eval.)
		16	Kurtosis, Skewness, etc + SVM	81.4	3	NO	8	43,200	YES	NO	YES (3 sbj. for training / 5 sbj. for eval.)
[12]	Pressure (FSR) + Video	60+video	PCA + SVM	94.1	9	YES	3	> 10,800	YES	NO	YES (*1)
[13]	Pressure Matrix	3,536	CoP, etc. + SVM	77.1	9	YES	10	270	YES	NO	YES (*1)
[14]	Pressure Matrix	3,200	CoP, etc. + SVM	92.3	4	YES	19	2,240	NO	YES	YES (*1)
			CoP, etc. + SVM	96.3	3	NO	19	2,240	NO	YES	YES (*1)
[15]	Pressure Matrix	2,048	PCA + kNN	97.7	5	NO	6	NA	YES	NO	NA
[16]	Pressure Matrix	2,048	Hamming distance similarity + kNN	97.1	8	NO	20	3,200	YES	NO	NA
[17]	Pressure Matrix	1,728	GMM + kNN	98.4	3	NO	9	NA	YES	YES	NA
[19]	Pressure Matrix	8,192	Sparse Classifier with Minimum Class Residual	83.2	6	YES	14	3,360	YES	NO	YES (*1)
[18]	Pressure Matrix	2,048	HOG+DNN	98.1	5	NO	10	NA+DataAug	YES	NA	NA
ours	Pressure Matrix	784 (subsampled from 3,200)	DNN	***99.7***	3	NO	19	224+DataAug	***NO***	***YES***	YES (*2, *3)
			DNN	***97.1***	4	***YES***	19	224+DataAug	***NO***	***YES***	YES (*2,*3)

*1: Leave one subject out cross validation, *2: Only one data for each posture, *3: Cross validation with subject based separation

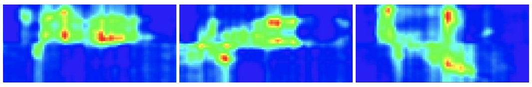

Fig. 13 Examples of prone posture data with estimation results w/ D.A.

teristics, and all data categorized correct have the same characteristics. Thus, the main factor of improvement about prone posture classification seems to be the position shift data augmentation.

6.3 Comparison with related works

The comparison with related works is summarized in **Table 11**. As shown in Table 11, our result has the highest accuracy in the group whose dataset includes a prone position. Our result also has the highest accuracy in the group whose dataset does not include a prone position, except [18]. The five postures of Ref. [18] are a supine position, two left lateral positions and two right lateral positions. These can be recalculated for the three postures discussed in this study, and then the recalculated accuracy of Ref. [18] becomes 100%. However, there are many things that are not discussed, such as base dataset size without data augmentation, training processes and so on. Our result showed almost 100% accuracy, 99.7%, with the described solid process. In addition, our DNN is simpler than ones of Ref. [18] and suitable for collaborations with related DNN techniques.

7. Conclusion

In this paper, we investigated appropriate data augmentation parameters for DNN based in-bed posture classification. The investigated parameters are ones of shear transformation, zoom, rotation, and shifts along short and long sides of a bed. As a result, the combination of the up to ±20% and ±40% random shifts along short and long sides of a bed, the up to ±10 degree random rotation, and the non-use of other transformations showed the best performance. With the parameters, the built DNN showed 99.7% accuracy and 0.997 Weighted F^1-score for three posture classifications: supine, left and right lateral positions, and 97.1% accuracy and 0.970 Weighted F^1-score for four posture classifications: supine, prone, left and right lateral positions. These results are the best performances among the compared related works in Section 6.3. In addition, our DNN described in this paper is simpler than those of related works and suitable for collaborations with related DNN techniques.

Acknowledgments This research is partly supported by JSPS KAKENHI Grant Number 15H02736, Strategic Information and Communications R&D Promotion Programme (SCOPE), and "The Knowledge Hub of AICHI, The Priority Research Project."

References

[1] Krizhevsky, A., Sutskever, I. and Hinton, G.E.: ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, pp.1097–1105 (2012).

[2] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A.: Going deeper with convolutions, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1–9 (2015).

[3] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778 (2016).

[4] Toshev, A. and Szegedy, C.: Deeppose: Human pose estimation via deep neural networks, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1653–1660 (2014).

[5] Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P.V. and Schiele, B.: Deepcut: Joint subset partition and labeling for multi person pose estimation, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.4929–4937 (2016).

[6] Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M. and Schiele, B.: Deepercut: A deeper, stronger, and faster multi-person pose estimation model, *European Conference on Computer Vision*, pp.34–50, Springer (2016).

[7] Wang, H., Lin, X., Zhang, Y. and Lee, T.S.: Learning Robust Object Recognition Using Composed Scenes from Generative Models, arXiv preprint arXiv:1705.07594 (2017).

[8] Sun, Z., Ozay, M. and Okatani, T.: Improving Robustness of Feature Representations to Image Deformations using Powered Convolution in CNNs, arXiv preprint arXiv:1707.07830 (2017).

[9] Lin, M., Chen, Q. and Yan, S.: Network in network, arXiv preprint arXiv:1312.4400 (2013).

[10] Xu, Y., Jia, R., Mou, L., Li, G., Chen, Y., Lu, Y. and Jin, Z.: Improved relation classification by deep recurrent neural networks with data augmentation, arXiv preprint arXiv:1601.03651 (2016).

[11] Hsia, C., Liou, K., Aung, A., Foo, V., Huang, W. and Biswas, J.: Analysis and comparison of sleeping posture classification methods using pressure sensitive bed system, *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp.6131–6134, IEEE (2009).

[12] Huang, W., Wai, A.A.P., Foo, S.F., Biswas, J., Hsia, C.-C. and Liou, K.: Multimodal sleeping posture classification, *2010 20th International Conference on Pattern Recognition* (*ICPR*), pp.4336–4339, IEEE (2010).

[13] Mineharu, A., Kuwahara, N. and Morimoto, K.: A study of automatic classification of sleeping position by a pressure-sensitive sensor, *2015 International Conference on Informatics, Electronics & Vision* (*ICIEV*), pp.1–5, IEEE (2015).

[14] Hayashi, C., Hara, K., Enokibori, Y. and Mase, K.: Sleeping Posture Classification Using E-Textile Pressure Sensor, *IPSJ SIG CVIM* (*Computer Vision and Image Media*) *Technical Reports*, Vol.2016, No.7, pp.1–6 (2016).

[15] Yousefi, R., Ostadabbas, S., Faezipour, M., Farshbaf, M., Nourani, M., Tamil, L. and Pompeo, M.: Bed posture classification for pressure ulcer prevention, *Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC 2011*), pp.7175–7178, IEEE (2011).

[16] Pouyan, M.B., Ostadabbas, S., Farshbaf, M., Yousefi, R., Nourani, M. and Pompeo, M.: Continuous eight-posture classification for bedbound patients, *6th International Conference on Biomedical Engineering and Informatics*, pp.121–126, IEEE (2013).

[17] Ostadabbas, S., Pouyan, M.B., Nourani, M. and Kehtarnavaz, N.: In-bed posture classification and limb identification, *2014 IEEE Biomedical Circuits and Systems Conference* (*BioCAS*), pp.133–136, IEEE (2014).

[18] Heydarzadeh, M., Nourani, M. and Ostadabbas, S.: In-bed posture classification using deep autoencoders, *the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (*EMBC 2016*), pp.3839–3842, IEEE (2016).

[19] Liu, J.J., Xu, W., Huang, M.-C., Alshurafa, N., Sarrafzadeh, M., Raut, N. and Yadegar, B.: Sleep posture analysis using a dense pressure sensitive bedsheet, *Pervasive and Mobile Computing*, Vol.10, pp.34–50 (2014).

Yu Enokibori received his B.E., M.E., and Ph.D. degrees in Engineering Science from Ritsumeikan University in 2005, 2007, and 2010, respectively. He became an assistant professor at Graduate School of Information Science, Nagoya University in 2015. He is now with Graduate School of Informatics, Nagoya University since 2017. His research interests include ubiquitous computing, wearable computing, invisible computing, healthcare and medical computing, and human computer interaction. He is a member of IPSJ, JSAI, SOBIM, JANS, and ACM.

Kenji Mase received his B.E. degree in Electrical Engineering and the M.E. and Ph.D. degrees in Information Engineering from Nagoya University in 1979, 1981 and 1992 respectively. He became a professor of Nagoya University in August 2002. He is now with the Graduate School of Informatics, Nagoya University. He is Research Supervisor of JST CREST on Symbiotic Interactions. He joined the Nippon Telegraph and Telephone Corporation (NTT) in 1981 and had been with the NTT Human Interface Laboratories until 1995. He was a visiting researcher at the Media Laboratory, MIT in 1988–1989. He has been with ATR (Advanced Telecommunications Research Institute) in 1995–2002. His research interests include gesture recognition, computer graphics, artificial intelligence and their applications for computer-aided communications, wearable/ubiquitous computers and lifelog. He is a Fellow of IEICE of Japan, and member of IPSJ, JSAI, Virtual Reality Society of Japan, Human Interface Society of Japan and ACM, and a senior member of IEEE Computer Society. He was a Section Chair of IEEE Nagoya Section in 2014–2015. He is an associate editor of ACM IMWUT journal and an editorial board member of the international journal on UMUAI.