Machine learning-based lunar calendar estimation using the lunar images

Yuto Ikenoue[1], Takashi Kuremoto[1, *], Shingo Mabu[2]
[1]Nippon Institute of Technology, Japan
[2]Yamaguchi University, Japan
kuremoto.takashi@nit.ac.jp

Abstract: The lunar calendar is estimated by the observation of the shape of the moon in the ancient time. In this study, we propose an ensemble machine learning method to determine the lunar calendar using the images of the moon. The proposed method uses multiple convolutional neural networks (CNNs) and the voting results of them. Experiment results showed that the proposed ensemble classifier was more effective than the conventional single CNN method.

Key-Words: *moon image, lunar calendar, CNN, ensemble learning, machine learning*

1. Introduction

In the past, people lived by the lunar calendar, which used the cycles of the phases of the moon. These lunar calendars use the cycle of the change of moon shapes from the new moon to the next new moon, whereas one cycle lasting approximately 29.5 days. Furthermore, the time elapsed from the new moon, expressed in days, is called the lunar age.

In modern times, the lunar calendar is used as the Islamic calendar in Islamic countries. However, in the Islamic calendar, the new moon does not refer to the moon that is in the same direction as the sun as seen from the earth, but to the narrow moon that becomes visible in the western sky in the evening again after the first day of the month. Therefore, it is said that the new moon is delayed by one to two days from the start of the real cycle of lunar month [1].

The lunar calendar is determined by observing the change the moon shapes (visible surface) and the estimation of the lunar age depends on the experience of the observers and the accuracy of the observations. In this study, we attempt to estimate the lunar age by machine learning methods using moon images.

Specifically, an ensemble learning method using 32 convolutional neural networks (CNNs) is proposed and compared to the conventional single CNNs. Experimental results by 36,000 lunar images confirmed the effectiveness of the proposed method.

2. Ensemble Learning

Ensemble learning is a machine learning method that integrates the prediction results from multiple weak discriminators to output more accurate prediction results. A case study is the validation of CIFAR-10 image classification by ensemble learning using CNNs by Sugihara et al. [2]. The experiment results suggested that the ensemble method had a higher recognition accuracy than single CNNs. The ensemble learning in this study accumulates the prediction probability results for 30 classes (lunar age) obtained from 32 single CNNs with different structures and outputs the class with the highest calculated value as the predicted month age.

Received: 2022/12/08, Accepted: 2023/04/21
*Corresponding author: Takashi Kuremoto
E-mail address: kuremoto.takashi@nit.ac.jp

3. Lunar Age Estimation Experiment

3.1 Building an ensemble learning model

In this study, a dataset of lunar surface images was created using an application called "Simple lunar age calendar" developed by Fumihiro Takahashi [3]. The image generation tool estimates the lunar age with dynamic lunar surface images. Since the average monthly age changes in a cycle of 29.5 days, the sample dataset of the moon images was created with 30 classification classes based on the number of days of lunar age. Detailly, class 1 is assigned to lunar ages between 0.0 and 1.0 days, class 2 assigned to lunar ages between 1.1 and 2.0 days, and so on. The 1,200 images in each class are named with a sequential number, the total number of the moon images in our dataset was 36,000. Two examples of these images are shown in Figure 1, where the moon areas were cropped in 100x100 squares in each image. Among 1,200 images of each class, 1,000 images were used as training samples, and 200 images were test (validation) samples.

class15_No.1 calss24_No.1

Figure 1. Two examples of moon images used in this study.

The weak discriminators were constructed using the Convolutional Neural Networks (CNNs), a deep learning model commonly used in image classification problem (Figure 2). The CNN is composed of three types of neural nets: convolutional layers, pooling layers, and Multi-Layered Perceptron (MLP) or so-called fully connection layers (FC). Additionally, it is optimized as a discriminator using training data and the error backpropagation (BP) learning rule.

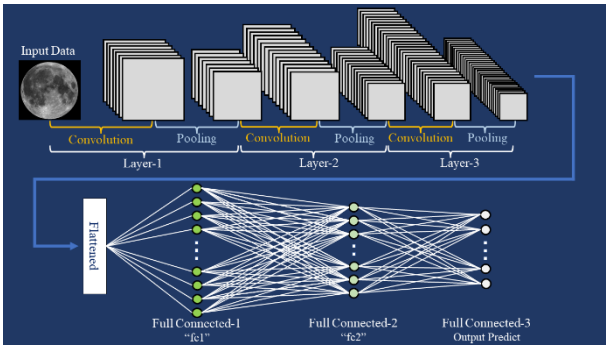


Figure 2. A structure of a single CNN used in this study.

In our experiment, 32 CNNs with the different structures were created to build an ensemble learner. The structures of each of the 32 CNNs are shown in Table 1. The Softmax function was used for the output layer as the activation function, while the ReLU function was used as units in each hidden layer of MLP. In Table 1, the values of “layer-1”, “layer-2”, and “layer-3” indicate the "number of convolution filters (kernel size of pooling)". The kernel size of all convolution filters is fixed to 3×3. “fc-1” and “fc-2” indicate the first layer and the second layer of MLPs in CNNs. The training time of each CNN was 100, learning rule was Adam optimizer [3].

$$t \leftarrow t + 1$$

$$g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$$

$$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

$$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{m}_t \leftarrow \frac{m_t}{(1 - \beta_1^t)}$$

$$\hat{m}_t \leftarrow \frac{m_t}{(1 - \beta_1^t)}$$

$$\theta_t \leftarrow \theta_{t-1} - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}$$

Where t is the training time, θ is the parameters of CNN, such as connection weights and biases, α is the learning rate usually 0.001. Other hyperparameters $\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8}$.

The ensemble learning model is finally composed by these single CNNs, i.e., the prediction probability

results for each class obtained from each CNN are accumulated and the class with the highest calculated value of the prediction probabilities is output as the final classification result.

Table. 1. Structural details of each CNN.

CNN	layer-1	layer-2	layer-3	fc-1	fc-2
1	64(5)	64(3)	64(2)	512	256
2	64(5)	64(3)	64(2)	1024	512
3	64(5)	64(2)	64(2)	512	256
4	64(5)	64(2)	64(2)	1024	512
5	64(4)	64(3)	64(2)	512	256
6	64(4)	64(3)	64(2)	1024	512
7	64(4)	64(2)	64(2)	512	256
8	64(4)	64(2)	64(2)	1024	512
9	128(5)	64(3)	64(2)	512	256
10	128(5)	64(3)	64(2)	1024	512
11	128(5)	64(2)	64(2)	512	256
12	128(5)	64(2)	64(2)	1024	512
13	128(4)	64(3)	64(2)	512	256
14	128(4)	64(3)	64(2)	1024	512
15	128(4)	64(2)	64(2)	512	256
16	128(4)	64(2)	64(2)	1024	512
17	128(5)	128(3)	64(2)	512	256
18	128(5)	128(3)	64(2)	1024	512
19	128(5)	128(2)	64(2)	512	256
20	128(5)	128(2)	64(2)	1024	512
21	128(4)	128(3)	64(2)	512	256
22	128(4)	128(3)	64(2)	1024	512
23	128(4)	128(2)	64(2)	512	256
24	128(4)	128(2)	64(2)	1024	512
25	128(5)	128(3)	128(2)	512	256
26	128(5)	128(3)	128(2)	1024	512
27	128(5)	128(2)	128(2)	512	256
28	128(5)	128(2)	128(2)	1024	512
29	128(4)	128(3)	128(2)	512	256
30	128(4)	128(3)	128(2)	1024	512
31	128(4)	128(2)	128(2)	512	256
32	128(4)	128(2)	128(2)	1024	512

3.2 Experiment Setup

The training of single CNNs was carried out using 30,000 training images and 6,000 validation images, and 5-fold cross-validation was conducted in the experiment. The data in each fold are shown in Table 2.

3.3 Estimation results

The change of estimation errors (losses) of the 5 splits data was shown in Figure 3, and the change of the classification accuracy was shown in Figure 4. It can be confirmed that the convergence of the errors and accuracies appeared after 50 epochs, i.e., CNN was optimized well.

Estimation results of the lunar calendar by the single CNNs and ensemble learning model are presented in Table 3. CNN-Ave is the average of the output results of each single CNN. *F1* represents the harmonic mean of *Precision* and *Recall*. As a result of the estimation, it was confirmed that ensemble learning model is prior to the conventional single CNN in all measures of performance evaluation.

Table 2. 5-fold cross-validation dataset.

	The number of images				
Split	1~20	201~40	401~60	601~80	801~100
1	Test	Train	Train	Train	Train
2	Train	Test	Train	Train	Train
3	Train	Train	Test	Train	Train
4	Train	Train	Train	Test	Train
5	Train	Train	Train	Train	Test

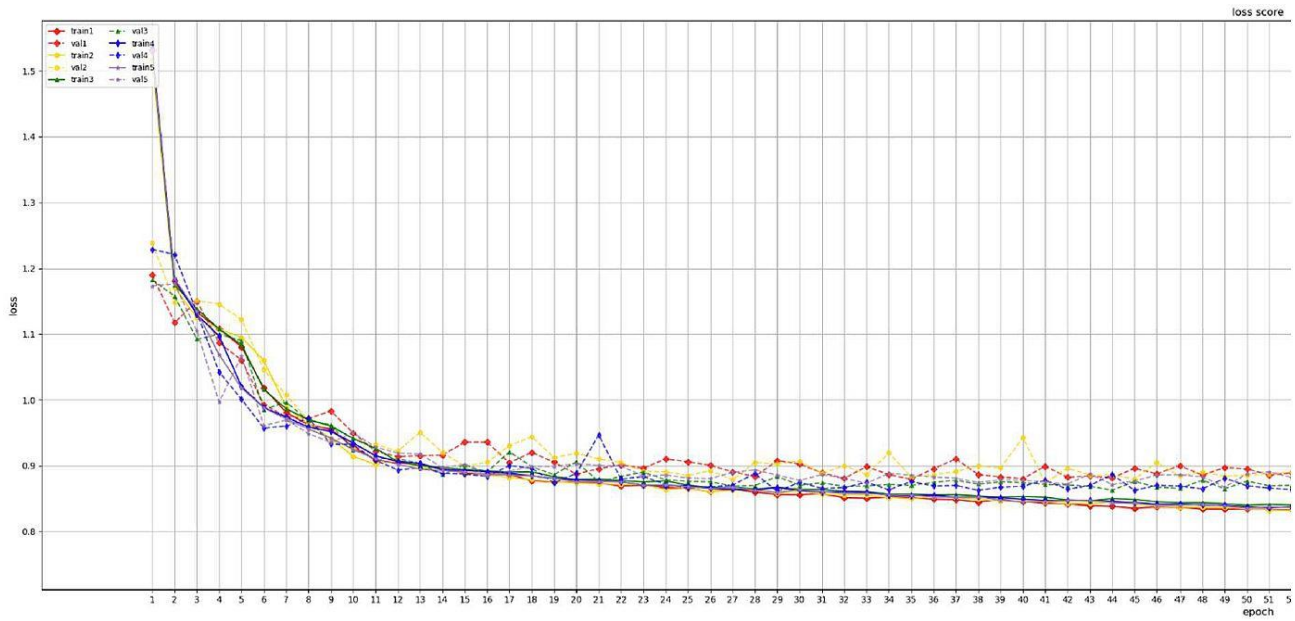


Figure 3. The change of loss according to the training time of CNN1.

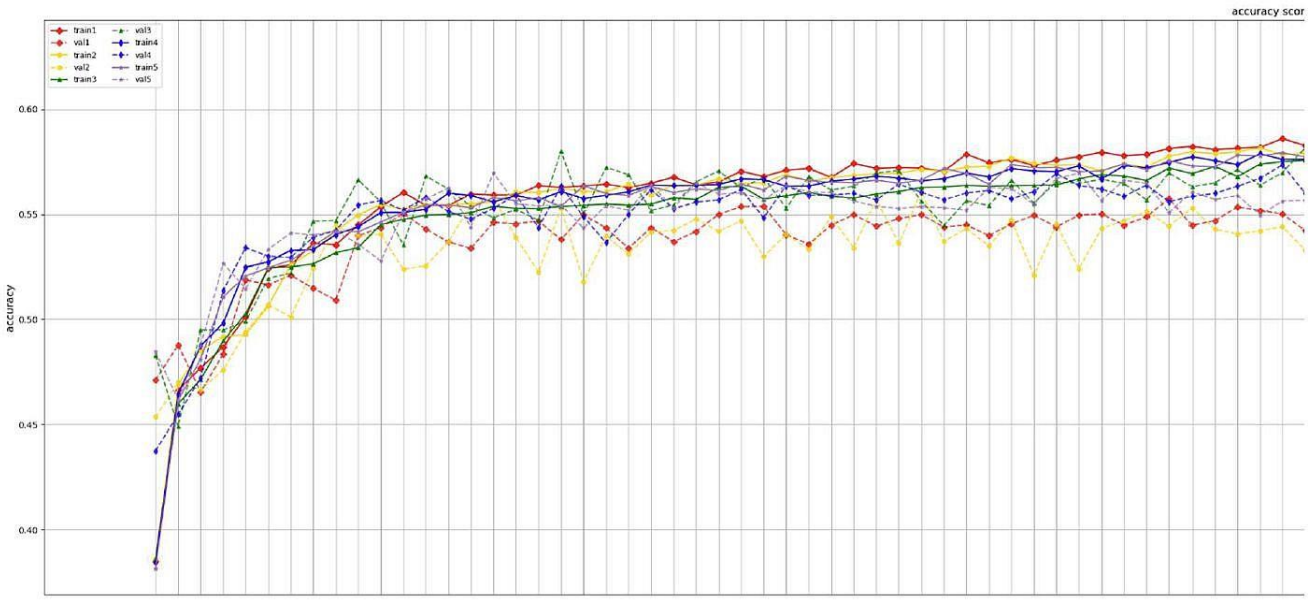
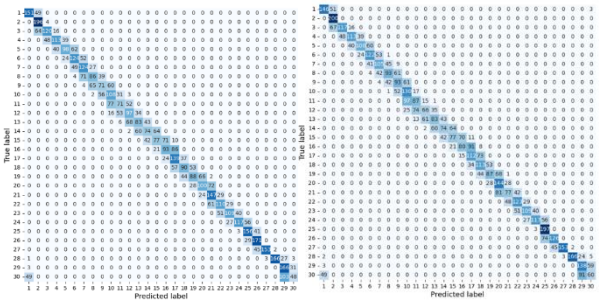


Figure 4. The change of accuracy according to the training time of CNN1.

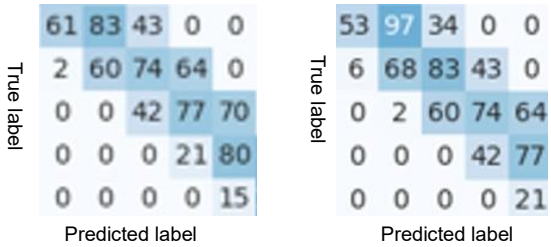
Table 3. Prediction of the single CNNs and ensemble learning model.

model	*Accuracy*	*Precision*	*Recall*	*F1*
CNN-Ave	0.5574	0.5632	0.5574	0.5407
Ensemble	**0.5639**	**0.5694**	**0.5639**	**0.5553**

Meanwhile, it can be found that not only CNN-Ave or ensemble learning model had a low prediction accuracy which was around 57%. To see how the models classified the lunar images, a confusion matrix was created as shown in Figure 5.



(a) Single CNN (b) Ensemble learning



(c) Enlarged image (CNN) (d)Enlarged image (Ensemble)

Figure 5. Single CNN and ensemble learning model confusion matrices

From Figure 5, it was confirmed that the predicted labels were concentrated on the correct classes which means these moon images had the similar shapes corresponding to the lunar ages.

Table 4. 5-fold cross-validation by an ensemble learning model.

Split	Accuracy	Precision	Recall	F1
1	0.5561	0.5608	0.5561	0.5503
2	0.5491	0.5542	0.5491	0.5370
3	0.5700	0.5667	0.5700	0.5563
4	0.5743	0.5849	0.5743	0.5680
5	0.5701	0.5806	0.5701	0.5651
Mean	**0.5639**	**0.5694**	**0.5639**	**0.5553**

4. Analysis

4.1 Analysis to cross-validation experiment results

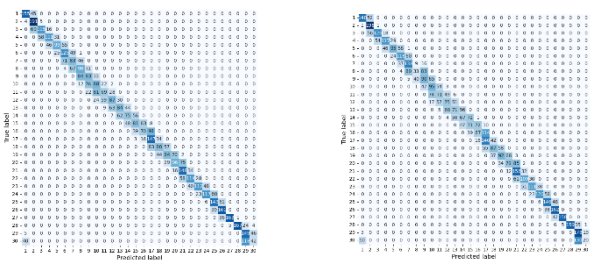
Using the datasets presented in Table 2, 5-fold cross-validation experiment for ensemble learning model performance evaluation was carried out, and the results are shown in Table 4. As a result of the cross-validation, the average accuracy 56.39% of the ensemble learning model is confirmed. The values were also close to the accuracy of ensemble learning model shown in Table 3.

Table 5. *t*-test using the results of 5-fold cross-validation for ensemble learning model

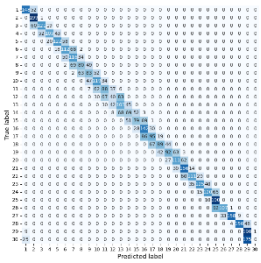
Split	Statistic	*p* value	$p<0.05$	$p<0.01$
1	-7.5677	1.56 × 10^(-8)	**Yes**	**Yes**
2	-2.7561	9.70 × 10^(-3)	**Yes**	**Yes**
3	-1.8436	7.48 × 10^(-2)	No	No
4	-11.5064	1.01 × 10^(-12)	**Yes**	**Yes**
5	-5.2106	1.17 × 10^(-5)	**Yes**	**Yes**

Data from each split obtained in the 5-fold cross-validation were analyzed by *t*-test of the ensemble learning model and 32 single CNNs and the result was shown in Table 5. The null hypothesis was set as “32 single CNNs and ensemble learning model have the same performance”, while the opposite hypothesis was set as “32 single CNNs and ensemble learning model do not have the same performance”. As a result, the conflict hypothesis was adopted in the cases $p<0.05$ and $p<0.01$ except Split 3.

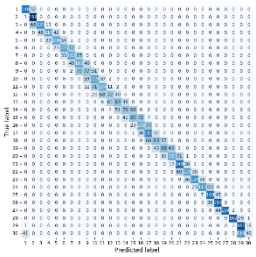
To see how the ensemble learning model classified the lunar images in the 5-fold cross-validation, a confusion matrix was shown in Figure 6. From the five confusion matrices Figure 6 (a) to Figure 6 (e), it can be confirmed that in class2, class25 to class29, the accuracies tend to be high. As same as the results shown in Figure 5, the predicted labels were concentrated on the correct classes which means these moon images had the similar shapes corresponding to the lunar ages.



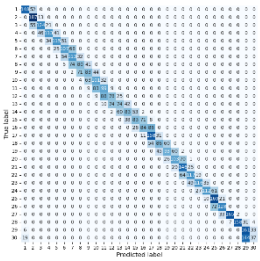
(a) Split-1 (b) Split-2



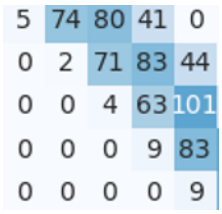
(c)plit-3



(d) Split-4



(e) Split-5



(f) Enlarged image (Split-5)

Figure 6. Confusion matrices of ensemble learning model in 5-fold cross-validation experiment.

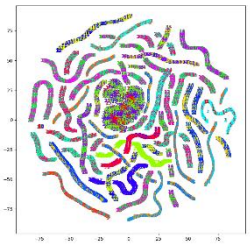
Table 6. Accuracies of single CNNs

CN	Split-	Split-	Split-	Split-	Split-	Mean
1	0.551	0.543	0.563	0.553	0.562	0.554
2	0.551	0.548	0.563	0.564	0.564	0.558
3	0.553	0.541	0.580	0.572	0.563	0.562
4	0.552	0.545	0.582	0.562	0.570	0.562
5	0.551	0.559	0.562	0.558	0.553	0.557
6	0.547	0.546	0.567	0.557	0.563	0.556
7	0.555	0.536	0.554	0.564	0.563	0.554
8	0.558	0.545	0.570	0.571	0.575	0.564
9	0.540	0.545	0.563	0.543	0.565	0.551
10	0.547	0.543	0.563	0.557	0.554	0.553
11	0.551	0.542	0.570	0.568	0.560	0.558
12	0.551	0.554	0.572	0.570	0.571	0.563
13	0.545	0.550	0.573	0.557	0.564	0.558
14	0.549	0.542	0.559	0.561	0.568	0.556
15	0.554	0.553	0.557	0.556	0.567	0.557
16	0.553	0.542	0.572	0.560	0.570	0.559
17	0.544	0.543	0.563	0.547	0.555	0.550
18	0.537	0.545	0.566	0.557	0.565	0.554
19	0.553	0.540	0.570	0.566	0.560	0.558
20	0.550	0.548	0.567	0.568	0.566	0.560
21	0.552	0.554	0.571	0.558	0.552	0.557
22	0.544	0.541	0.558	0.561	0.546	0.550
23	0.553	0.544	0.558	0.558	0.569	0.556
24	0.551	0.549	0.573	0.563	0.568	0.561
25	0.547	0.544	0.567	0.560	0.567	0.557
26	0.541	0.541	0.552	0.564	0.558	0.551
27	0.551	0.541	0.583	0.565	0.577	0.563
28	0.547	0.550	0.567	0.559	0.569	0.558
29	0.542	0.549	0.560	0.555	0.552	0.551
30	0.550	0.549	0.573	0.557	0.563	0.558
31	0.553	0.556	0.571	0.571	0.569	**0.564**
32	0.536	0.549	0.576	0.558	0.556	0.555

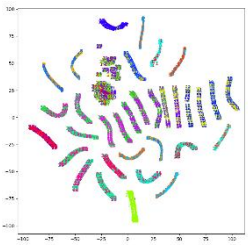
The accuracies of 32 single CNNs in 5-fold cross-validation experiment is shown in Table. 6 The maximum value was 56.42% given by CNN 31 and all CNNs had accuracies more than 55.0%. The mean value was 55.74%, which is close to the values of them calculated in Table 3.

4.2 Visualization of datasets using t-SNE

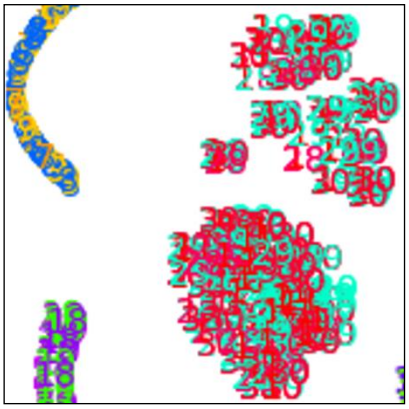
Tables 3 and 4 show that the accuracies of single CNNs and the proposed ensemble learning model are around 56%, and from Figures 2 and 3, we considered that there were some input images close to each other, so it reduced the accuracies of classification. To observe the distribution of the dataset, "t-SNE", proposed by Maaten and Hinton et al. [5] was adopted to transform these high-dimensional data (images) into a two-dimensional space. The transforming results are shown in Figure 7.



(a) Training data



(b) Validation data



(c) Enlarged image (Validation data)

Figure 7. Visualization of datasets using t-SNE [5].

Figure 7 shows a color-coded plot of each class in the moon image dataset. The plotted points are densely clustered together with those that are close to each other in lunar age Figure 7 (c). This is thought to be due to the slight differences between the beginning and ending of the lunar cycle. For this nature of the lunar age, the average accuracy of estimation for each day were not high, but around 57%.

4.3 Ensemble mode with more CNNs

The number of CNNs in the ensemble model of Section 3 was 32, which was decided by the consideration of problem of 30 classes estimation. We also performed an experiment with 50 CNNs, and the results showed the average accuracy of estimation was improved from 56.39% to 56.50%, and Precision, Recall, F1 values were also raised. Detail results were shown in Table 7 and 8.

Table 7 5-fold cross-validation results by an ensemble model with 50 CNNs (%)

Split	Accuracy	Precision	Recall	F1
Split-1	55.81	56.36	55.81	55.12
Split-2	54.83	55.43	54.83	54.05
Split-3	57.15	59.77	57.15	56.17
Split-4	57.38	58.38	57.38	56.86
Split-5	57.35	58.53	57.35	56.85
Mean (SD)	56.50 (1.01)	57.69 (1.57)	56.50 (1.01)	55.81(1.08)

Table 8 Comparison of estimation results by different models.

Model	Accuracy	Precision	Recall	F1
CNN-Ave	55.92	56.45	55.94	54.7
32 CNNs	56.39	56.94	56.39	55.53
50 CNNs	**56.50**	**57.69**	**56.50**	**55.81**

4.4 Top1-2-3 prediction accuracies

In the fundamental view on the nature of the lunar age analyzed in Section 4.2, it led us to consider that the predictions of Top1-2-3 (correctness of the output of models is allow up to 3 candidates) may relatively high. The predict accuracies of Top1, Top1-2, and Top1-2-3 by ensemble learning models are shown in Table 9. It can be confirmed that the accuracy of Top1-2-3 marked a high level, i.e., 99.62% by the ensemble model with 32 CNNs and 99.58% by the ensemble model with 50 CNNs.

Table 9 Mean accuracies in different measures (%).

	Top1 (SD)	Top1-2 (SD)	Top1-2-3 (SD)
32 CNNs	56.39 (0.96)	90.62 (0.37)	**99.62 (0.09)**
50 CNNs	**56.50 (1.01)**	**90.67 (0.37)**	99.58 (0.13)

Table 10 Computation time of different models for Split-1 data (sec/100 images).

	Mean	SD
Single CNNs with CPU	2.13	0.58
Single CNNs with GPU	0.47	0.25
50 CNNs with CPU	68.19	-
50 CNNs with GPU	15.08	-

4.5 Comparison of computation time

The training time of different models were compared in Table 10. It was shown that bigger models needed more

time and the computer with GPU had a double speed to the case of CPU only computer. The computer with CPU only was Intel Core i-7-9700k, 32GB ROM, and the computer with GPU used Geforce RTX 2080SUPER.

5. Conclusion

In this study, an ensemble learning model composed of multiple CNNs with different structures was proposed for lunar calendar estimation using moon images. Experiment results showed that the proposed method raised the estimation accuracy rate which was higher than using single CNNs. Future work may include improving the performance of the weak discriminator, by adopting more different advanced CNNs such as ResNet50, EfficientNet, YOLOv5, and the hybrid model, e.g., CNN with SVM. Estimation using real images or an online practical application development also remains in the future.

Acknowledgement:

This work was supported by JSPS KAKENHI Grant No.22H03709, and No. 22K12152.

References:

[1] National Astronomical Observatory of Japan Calendar Computing Office, “Lunar calendar” (in Japanese), https://eco.mtk.nao.ac.jp/koyomi/wiki/C2C0B1A2CEF1.html

[2] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, "A survey on ensemble learning," Frontiers of Computer Science, vol. 14, no. 2, pp. 241258, 2020

[3] D. P. Kingma, J. Ba, “Adam: A method for stochastic optimization”, arXiv:1412.6980, 2017

[4] Fumihiro Takahashi. “Simple Lunar Calendar” (in Japanese), https://apps.apple.com/jp/app/id1279710898

[5] Laurens van der Maaten and Geoffrey Hinton, “Visualizing Data using t-SNE”, Journal of Machine Learning Research, Vol. 9, Issue 86, pp. 2579-2605, 2008