# ANALYSIS OF LUNAR SURFACE DATA USING MACHINE LEARNING: IDENTIFICATION OF SUNLIT AREA AND SHADE AREA OF HIGH LATITUDE AREA USING KAGUYA SP DATA.

Haramitsu Kyoda (hkyoda@planeta.sci.isas.jaxa.jp) [1,2], Yoshiaki Ishihara[2] , Satoshi Tanaka[2], Hiroka Inoue[2], Makiko Ohtake[2], Hisashi Otake[2]. [1]Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwashi, Chiba, 277-8562, Japan. [2]Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Chuou-Ku, Sagamihara, Kanagawa 252-5210, Japan.

**Introduction:** Recently, studies of lunar polar region is getting popular for research of volatile elements and future landing site selection including manned missions [e.g. [1], [2]].

The lunar explorer Kaguya(SELENE) acquired a large mass of information both topographical and reflection spectrum using Terrain Camera(TC), Multi band Imager(MI), and Spectral Profiler(SP) [3] and so on. Among them, the SP data of visible to near infrared spectrum with high S/N ratio, which was obtained more than 10 million shots at the latitude higher than 80 degrees, contains useful mineralogical information. In practical, all the SP data is usually corrected under the assumption of sunlit condition, which may not be affected so much taken in the low latitude area. However, accurate data correction by sunshine conditions (including secondary reflected) should be done in order to analyze the data properly in the high latitude area. Currently, only spectral with very high S/N ratio in the sunlit area are used for analysis. However, such data is rare in the polar region. For this reason, there are many data remained to be analyzed. Also, "Half-shaded" area at where only the secondary reflected light is exposed should be focused to search for volatile elements [1]. Therefore, In this study, we developed a automatic classifier of sunlit and shade area using a machine learning method.

**Method:** SP data is a point spectrum, but MI or TC images were taken almost simultaneously with SP observations. These data (so called SP support data) which are attached to the SP profile data, are conventionally used to identify the interested location to be analyzed. Because images in the high latitude area are much darker than those of low latitude data, it is difficult to distinguish between secondary reflected light and direct sunlight. In order to solve this, we conducted ray-tracing simulation to of a compatible area of Digital Elevation Model(DEM)[4] under the appropriate condition (Kaguya orbit, lunar ephemeris and solar position) at the timing of SP data obtained. However, it is difficult to simulate the whole area of the Moon because the exact orbit data of the Kaguya are limited Images obtained through the simulation using DEM were given as teacher data. And we attempted to develop an automatic classifier which can recognize the sunlit area and shade area. In this study, we adopted the algorithm called Fully Convolutional Network (FCN) [5] which is commonly used for Automatic driving technique, and 23 pairs of simulated and SP support images were used as teacher data. We used the divided data of SP support images and simulated images into pixel size of 128 x 128 for deep learning (Fig.1). A relationship between loss and epoch is shown in Fig.2. We considered that there is no problem if the number of epochs is 25 or more.

**Results and Discussion:** In order to verify the validity of the data processing, binary images of sunlit area and shade area were calculated from three SP support images at the latitude of north latitude 73.2 degrees, North latitude 77.1 degrees, and Southern latitude 88.4 degrees respectively. For comparison, ray-tracing simulation was also performed at the same areas. The results applied to the 3 regions are shown in Table.1. The overall percentage of correct answer rate was about 98%(Table.1). It would be practically sufficient to determine whether the observation point is sunlit area or not. When we compare the result in detail shown in Fig.3 and Fig.4, green area which indicates falsely detected as a sunlit area, is typically shown at Fig.3. On the other hand, and blue area which indicates falsely detected as a shade area, is typically shown at right of fig4. In addition, we found that the false detection rate is higher in the vicinity of the boundary between the shade area and the sunlit area. This is thought to be caused by the resolution drops at the pooling layer of FCN[5]. As a whole, global features can be successfully reproduced as long as data correction.

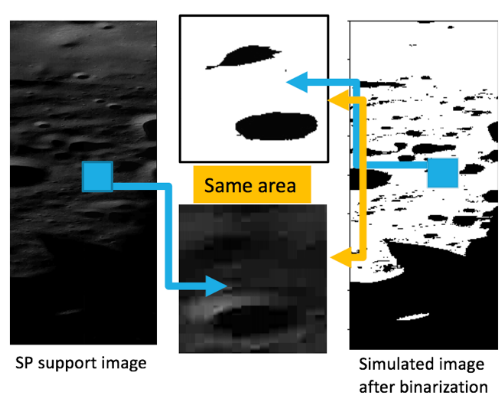


**Fig.1:***(left):SP support image. (Right):An image in which a simulated image is binarized into a sunlit area and a shaded area. (Center):128 x 128pixel divided images inputted the deep learning.*

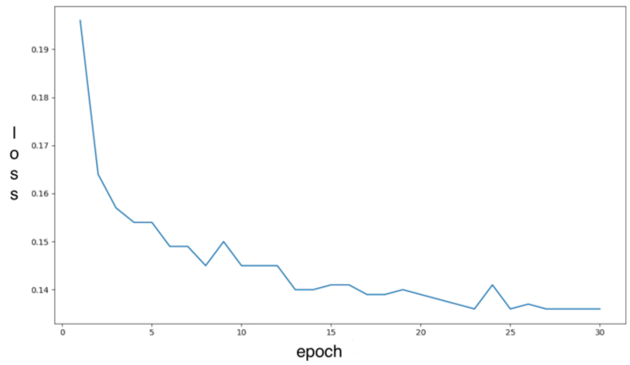


**Fig.2:** *A relationship between error loss and number of learning (epoch).*

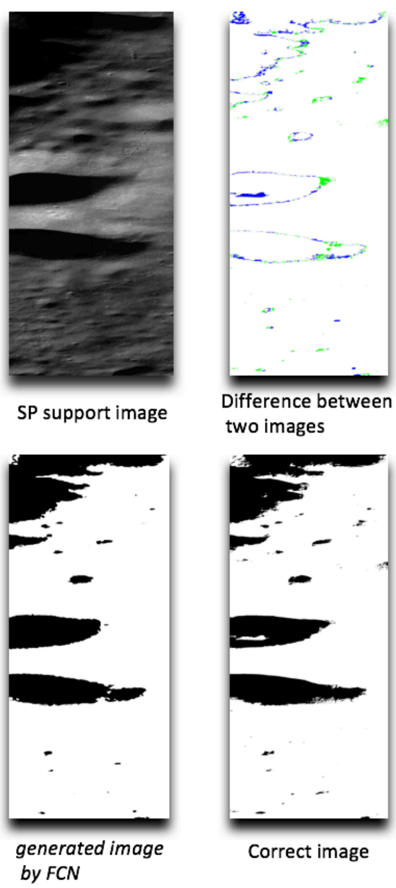


**Fig.3:** *(Upper left): Input image to put in the classifier (SP Support image at around Eastern longitude 143.5 degrees North latitude 77.1 degrees.) (Upper right): An image showing the difference between the correct image and the generated image by FCN. The green area is mis-identified as shaded area the blue area is mis-identified as the sunlit area. (Under left): Correct image(An image obtained by binarizing a simulated image). (Under right): generated image by FCN.*

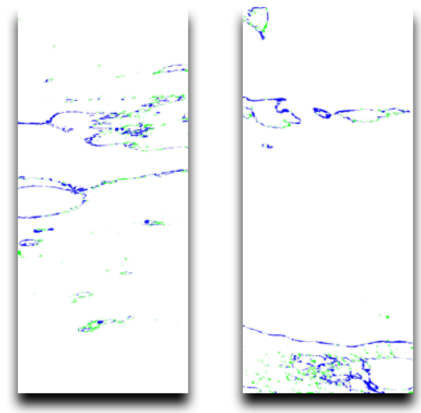

***Fig.4:*** *(Left): Difference between the correct image in the vicinity of Eastern longitude 140.6 degrees north latitude 73.2 degrees and the generated image by FCN(See Figure 3 for details). (Right): Difference between the correct image in the vicinity of Eastern longitude 115.7 degrees Southern latitude 88.1 degrees and the generated image by FCN.*

**Table.1:** *Percentage of Correct answer rate for three verification images. Sunlit Accuracy shows the correct answer rate in the sunlit area and Shadow Accuracy shows the correct answer rate in the shade area.*

| filename | Sunshine area rate | Shadow area rate | Sunshine accuracy | Shadow accuracy | Total accuracy |
|---|---|---|---|---|---|
| N732 | 0.834 | 0.166 | 0.992 | 0.924 | 0.981 |
| N771 | 0.812 | 0.188 | 0.987 | 0.929 | 0.976 |
| S884 | 0.122 | 0.878 | 0.947 | 0.983 | 0.979 |

**Summary:** By machine learning, we succeeded in discriminating sunlit area and shade area with accuracy of 97% or more.-applying other machine-learning methods such as Pyramid Scene Parsing Network are considered to make more accurate classifier.

**References:** [1] Heldmann, Jennifer L., et al. (2016) Acta Astronautica, 127, 308-320. [2] Ivanov, M. A., et al. (2015) Planetary and Space Science, 117, 45-63. [3] Yamamoto, S., et al. (2011) IEEE, 49, 4660-4676. [4] Honda, R. et al. (2008), LPSC 39, #1876. [5] Jonathan, L., et al. (2015) arXiv, 1411.4038v2, 1-10.