Aerosol Optical Depth Retrieval for Sentinel-2 Based on Convolutional Neural Network Method

Jie Jiang 1,2,*, Jiaxin Liu 1 and Donglai Jiao 1,2

1 Department of Surveying and Geoinformatics, Nanjing University of Posts and Telecommunications, Nanjing 210023, China; 1021173510@njupt.edu.cn (J.L.); jiaodonglai@njupt.edu.cn (D.J.)
2 Smart Health Big Data Analysis and Location Services Engineering Lab. of Jiangsu Province, Nanjing 210023, China
* Correspondence: jiangj@njupt.edu.cn

Abstract: Atmospheric aerosol significantly affects the climate environment and public health, and Aerosol Optical Depth (AOD) is a fundamental optical characteristic parameter of aerosols, so it is important to develop methods for obtaining AOD. In this work, a novel AOD retrieval algorithm based on a Convolutional Neural Network (CNN) method that could provide continuous and detailed aerosol distribution is proposed. The algorithm utilizes data from Sentinel-2 and Aerosol Robotic Network (AERONET) spanning from 2016 to 2022. The CNN AOD data are consistent with the AERONET measurements, with an $R^2$ of 0.95 and RMSE of 0.049 on the test dataset. CNN demonstrates superior performance in retrieving AOD compared with other algorithms. CNN retrieves AOD well on high reflectance surfaces, such as urban and bare soil, with RMSEs of 0.051 and 0.042, respectively. CNN efficiently retrieves AOD in different seasons, but it performs better in summer and winter than in spring and autumn. In addition, to study the relationship between image size and model retrieval performance, image datasets of 32$\times$32, 64$\times$64 and 128$\times$128 pixels were created to train and test the CNN model. The results show that the 128-size CNN performs better because large images contain rich aerosol information.

Keywords: AOD; CNN; Sentinel-2; deep learning

1. Introduction

Atmospheric aerosols are solid, liquid, or solid–liquid mixed particles suspended in the atmosphere with an aerodynamic diameter of less than 100 µm. These particles are mainly distributed in the stratosphere and troposphere at the lower portion of the atmosphere. Aerosols exhibit a complex and variable composition resulting from a combination of natural emissions and human activities, including anthropogenic aerosols, such as sulfates and biomass-burning aerosols, as well as natural aerosols, such as dust and sea salt, mineral dust, and primary biological aerosol particles [1].

Aerosol, which plays a crucial role in the Earth’s atmospheric system, has a significant influence on radiation balance, climate change, atmospheric environmental quality, and human health [2–6]. Aerosol particles could change Earth’s radiation budget through scattering and absorbing incoming solar radiation [7]. Aerosols also have an indirect effect on climate by altering cloud microstructure [8,9]. In addition to the above-mentioned aerosol cloud microphysical effects, aerosols can also change the atmospheric thermodynamic characteristics through aerosol–radiation interaction, thereby altering the atmospheric circulation, thus affecting the occurrence of clouds and precipitation [10]. For example, black carbon aerosols can heat the atmosphere, evaporate cloud droplets to reduce cloud cover [11], create stable atmospheric stratification [12], weaken convection, and reduce rain showers. High concentrations of aerosol particles seriously affect human health [13–15]. A number of studies have shown that the incidence of cardiopulmonary disease is closely related to the content of particulate matter in polluted air [16–18], especially the inhalable...
fine particles, such as PM$_{2.5}$, which can be inhaled into the lungs, attached to the alveoli, or entered into the blood circulation through the alveoli, posing a great threat to human health [19,20]. Outdoor air pollution, primarily caused by PM$_{2.5}$, results in the premature deaths of tens of thousands of people worldwide every year [21]. Moreover, the chemical composition of aerosol particles may contain or adsorb toxic gases, viruses, bacteria, and heavy metal elements in the air, which will increase the risk of cancer after entering the human body [22]. Accordingly, the need for monitoring and controlling aerosols has become increasingly crucial. Aerosol Optical Depth (AOD) is a fundamental optical characteristic parameter that is the integral of the extinction coefficient of the atmospheric medium in the vertical direction. AOD is a vital indicator for monitoring aerosols and an indispensable parameter for estimating PM$_{2.5}$. Therefore, it is important to obtain accurate AOD.

At present, detection instruments, such as multi-band sun-photometers and lidar, are widely used to obtain aerosol property parameters. The accuracy of AOD data provided by the ground-based aerosol network is high. However, continuous spatiotemporal data are difficult to provide due to the limitations of high operating costs, uneven site distribution, and limited spatial coverage. Satellite remote sensing has emerged as an important means for aerosol monitoring and research due to its ability to provide a rapid response and large-area, long-term series monitoring. In the past few decades, several scholars have exerted efforts to explore the properties of aerosols from satellite images and have accomplished numerous achievements [23–26]. The famous Dark Target (DT) algorithm is a widely used aerosol retrieval algorithm that uses the linear relationship between the visible band and mid-infrared band to estimate the surface reflectance of red and blue bands and retrieves AOD from Moderate Resolution Imaging Spectroradiometer (MODIS) images according to the lookup table [27]. The DT algorithm is highly effective in retrieving AOD and can be used to retrieve AOD on vegetation surface and ocean areas. Nevertheless, this algorithm may not be applicable in desert and snowy regions. Considering the difficulty of utilizing instruments to retrieve aerosols in arid, semi-arid, and urban areas, some scholars developed the Deep Blue (DB) algorithm, which assumes that the surface reflectance of most figures remains unchanged in a relatively short period of time and retrieves AOD by creating a seasonal surface reflectance database [28]. Subsequently, a highly accurate Multi-angle Implementation of Atmospheric Correction (MAIAC) algorithm was developed on the basis of the dynamic time series analysis [29]. This algorithm can be used to retrieve the surface bidirectional reflection factor and AOD on vegetation surface and bright targets [30]. In recent years, people have paid more attention to air pollution, resulting in increased demand for high-accuracy aerosol retrieval, and the emergence of deep learning has brought new options for high-accuracy aerosol retrieval. Deep learning has strong nonlinear fitting ability [31,32] that can discover the potential behavior of the system from long-term observations [33]. This mechanism can be used as a statistical method to extract remote sensing information for quantitative parameter retrieval problems. Chen et al. [34] retrieved AOD in China based on Himawari-8 satellite data and deep neural networks, with a correlation coefficient of up to 0.87. Chen et al. [35] developed an aerosol retrieval model based on Convolutional Neural Network (CNN), and 68% of the retrieved AOD values are within the MODIS expected error (EE) envelope over land of $\pm (0.05 + 15\%)$. In-depth research on deep learning by scholars shows that deep learning has the superiority of high accuracy in predicting atmospheric parameters [36–39].

In this paper, a new aerosol retrieval method based on CNN was developed and used Sentinel-2 satellite images from 2016 to 2022 to retrieve AOD. CNN could directly retrieve AOD from satellite images, simplifying the steps of aerosol retrieval and improving its stability and spatiotemporal adaptability. CNN could provide continuous and detailed aerosol distribution to fill the observation gap in the ground-based monitoring network and provide a theoretical basis and application method for the research on climate change and environmental pollution. Section 2 describes the satellite images and ground-based measurement AOD data for the CNN model training and testing. Section 3 gives the details of data processing and CNN model structure. The validation results of the CNN AOD and
the CNN performance from different seasons, surface types, and scales are discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2. Materials

2.1. Aerosol Robotic Network (AERONET) Data

AERONET is a global aerosol optical property monitoring network established by the National Aeronautics and Space Administration. This network uses the CIMEL automatic sun-photometer as the basic observing instrument [40], which can observe aerosol characteristics at eight wavelengths from visible to near-infrared every 15 min, and the network consists of more than 1500 sites around the world [41]. These sites have been providing researchers with long-term, stable, and easily accessible aerosol data for decades. AERONET aerosol data are divided into three levels: Level 1.0 (unscreened), Level 1.5 (cloud-cleared and quality controls), and Level 2.0 (automatically cloud-cleared and quality-assured). The number of ground-based AODs available at level 2.0 is limited. Accordingly, level 1.5 data are selected for unified sampling to have sufficient samples. East Asia is the ideal research area for training and testing an AOD retrieval model because of the complex and variable composition of aerosols and abundant distribution of AERONET sites. V3 Level 1.5 data from 22 sites in East Asia from 2016 to 2022 were obtained in this study based on the climate, the land cover types around the sites, and the effectiveness and availability of observation data. Figure 1 shows the distribution of 22 sites used in this work, and Table 1 provides the site name, location, abbreviation, corresponding coordinates, and matched number with satellite images of each site.

![Figure 1. Locations of the AERONET sites used in this study.](image)

Table 1. Ground-based monitoring sites information.

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Location</th>
<th>Abbreviation</th>
<th>Longitude (° E)</th>
<th>Latitude (° N)</th>
<th>Matched Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>North China</td>
<td>NC01</td>
<td>116.381</td>
<td>39.977</td>
<td>289</td>
</tr>
<tr>
<td>Beijing-CAMS</td>
<td>North China</td>
<td>NC02</td>
<td>116.317</td>
<td>39.933</td>
<td>311</td>
</tr>
<tr>
<td>Beijing_PKU</td>
<td>North China</td>
<td>NC03</td>
<td>116.310</td>
<td>39.992</td>
<td>170</td>
</tr>
<tr>
<td>Beijing_RADI</td>
<td>North China</td>
<td>NC04</td>
<td>116.379</td>
<td>40.005</td>
<td>181</td>
</tr>
<tr>
<td>AOE_Baotou</td>
<td>North China</td>
<td>NC05</td>
<td>109.629</td>
<td>40.852</td>
<td>202</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Site Name</th>
<th>Location</th>
<th>Abbreviation</th>
<th>Longitude (° E)</th>
<th>Latitude (° N)</th>
<th>Matched Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hong_Kong_PolyU</td>
<td>Hong Kong</td>
<td>HK01</td>
<td>114.180</td>
<td>22.303</td>
<td>63</td>
</tr>
<tr>
<td>Hong_Kong_Sheung</td>
<td>Hong Kong</td>
<td>HK02</td>
<td>114.117</td>
<td>22.483</td>
<td>25</td>
</tr>
<tr>
<td>NAM_CO</td>
<td>Tibet</td>
<td>TB01</td>
<td>90.962</td>
<td>30.773</td>
<td>18</td>
</tr>
<tr>
<td>QOMS_CAS</td>
<td>Tibet</td>
<td>TB02</td>
<td>86.948</td>
<td>28.365</td>
<td>90</td>
</tr>
<tr>
<td>Lulin</td>
<td>Taiwan</td>
<td>TW01</td>
<td>120.874</td>
<td>23.469</td>
<td>90</td>
</tr>
<tr>
<td>Chen-Kung_ Univ</td>
<td>Taiwan</td>
<td>TW02</td>
<td>120.205</td>
<td>22.993</td>
<td>75</td>
</tr>
<tr>
<td>Douliu</td>
<td>Taiwan</td>
<td>TW03</td>
<td>120.545</td>
<td>23.712</td>
<td>55</td>
</tr>
<tr>
<td>EPA-NCU</td>
<td>Taiwan</td>
<td>TW04</td>
<td>121.185</td>
<td>24.968</td>
<td>84</td>
</tr>
<tr>
<td>Taipei_CWB</td>
<td>Taiwan</td>
<td>TW05</td>
<td>121.538</td>
<td>25.015</td>
<td>52</td>
</tr>
<tr>
<td>Chiba_University</td>
<td>Japan</td>
<td>JP1</td>
<td>140.104</td>
<td>35.625</td>
<td>132</td>
</tr>
<tr>
<td>Hokkaido_University</td>
<td>Japan</td>
<td>JP3</td>
<td>141.341</td>
<td>43.076</td>
<td>63</td>
</tr>
<tr>
<td>Noto</td>
<td>Japan</td>
<td>JP4</td>
<td>137.137</td>
<td>37.334</td>
<td>52</td>
</tr>
<tr>
<td>Baengnyeong</td>
<td>South Korea</td>
<td>KR1</td>
<td>124.630</td>
<td>37.966</td>
<td>15</td>
</tr>
<tr>
<td>Gangneung_WNU</td>
<td>South Korea</td>
<td>KR2</td>
<td>128.867</td>
<td>37.771</td>
<td>309</td>
</tr>
<tr>
<td>Gwangju_GIST</td>
<td>South Korea</td>
<td>KR3</td>
<td>126.843</td>
<td>35.228</td>
<td>70</td>
</tr>
<tr>
<td>Hankuk_UFS</td>
<td>South Korea</td>
<td>KR4</td>
<td>127.266</td>
<td>37.339</td>
<td>141</td>
</tr>
</tbody>
</table>

2.2. Sentinel-2 Data

Sentinel-2 is an Earth observation mission in the European Space Agency’s Copernicus program, and it consists of two polar-orbiting satellites: Sentinel-2A and Sentinel-2B. Every satellite carries a multi-spectral instrument to acquire 13 spectral bands from visible to shortwave infrared along a 290 km-wide orbital belt. The revisit period is 10 days for each satellite and 5 days for the combined system [42]. Sentinel-2 provides land surface images with spatial resolutions of 10, 20, and 60 m/pixel covering the global land, which is convenient for scholars to apply in remote sensing analysis fields, such as forest monitoring, land cover change monitoring, and natural disaster management. Sentinel-2 is superior for characterizing high spatial resolution and relatively high temporal resolution simultaneously, making it possible to obtain a continuous and detailed aerosol distribution map. The European Space Agency provides free Sentinel-2 Level-1C products, which are the atmospheric apparent reflectance products that have been finely orthorectified and sub-pixel geometric corrected. In this study, Sentinel-2 L1C images from Bands 2, 4, 8, and 12 from 2016 to 2022 were selected (Table 2).

Table 2. Information of Sentinel-2 data used in this study.

<table>
<thead>
<tr>
<th>Bands</th>
<th>Spatial Resolution (m/pixel)</th>
<th>Central Wavelength (nm)</th>
<th>Bandwidth (nm)</th>
<th>Central Wavelength (nm)</th>
<th>Bandwidth (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 2</td>
<td>10</td>
<td>492.4</td>
<td>66</td>
<td>492.1</td>
<td>66</td>
</tr>
<tr>
<td>Band 4</td>
<td>10</td>
<td>664.6</td>
<td>31</td>
<td>664.9</td>
<td>31</td>
</tr>
<tr>
<td>Band 8</td>
<td>10</td>
<td>832.8</td>
<td>106</td>
<td>832.9</td>
<td>106</td>
</tr>
<tr>
<td>Band 12</td>
<td>20</td>
<td>2202.4</td>
<td>175</td>
<td>2185.7</td>
<td>185</td>
</tr>
</tbody>
</table>

2.3. MODIS and Himawari-8 Data

MODIS has provided scholars with different types of global aerosol products for decades, among which MODIS AOD standard products over land include AOD retrieved by the MAIAC algorithm, DT algorithm, DB algorithm, and combined Dark Target and Deep Blue (DTB) [34]. At present, the MODIS MAIAC aerosol products have been developed for the C6.0 version, and the DT, DB, and DTB aerosol products have been developed for the C6.1 version [28]. In 2015, the Japan Meteorological Agency began operating the geostationary satellite Himawari-8 and providing remote sensing data with a high temporal resolution (10 min) [43]. The MODIS data used in this study are the daily aerosol
product MCD19A2 of C6.0 and MOD04_L2 of C6.1 Level 2 with spatial resolutions of 1 and 10 km/pixel, respectively (Table 3). The Himawari-8/AHI level 3 hourly Yonsei Aerosol Retrieval (YAER) product with a spatial resolution of 0.05°/pixel was used in this work [44]. These aerosol products were selected to evaluate the reasonability of the spatial distribution of the CNN AOD.

Table 3. Information of MODIS and Himawari-8 data used in this study.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Satellite/Sensor</th>
<th>Contents</th>
<th>Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCD19A2</td>
<td>Terra/MODIS</td>
<td>MAIAC AODs</td>
<td>1 km/pixel</td>
</tr>
<tr>
<td>MOD04_L2</td>
<td>Terra/MODIS</td>
<td>DT AODs</td>
<td>10 km/pixel</td>
</tr>
<tr>
<td></td>
<td>Terra/MODIS</td>
<td>DB AODs</td>
<td>10 km/pixel</td>
</tr>
<tr>
<td></td>
<td>Terra/MODIS</td>
<td>DTB AODs</td>
<td>10 km/pixel</td>
</tr>
<tr>
<td>H08</td>
<td>Himawari-8/AHI</td>
<td>YAER AODs</td>
<td>0.05°/pixel</td>
</tr>
</tbody>
</table>

3. Methods

3.1. Data Preprocessing

The dataset was constructed using multiple data preprocessing steps, including band interpolation, resampling, pixel screening, and spatiotemporal matching (Figure 2). Considering that CNN AOD must be compared with MODIS aerosol products, MODIS AOD is the AOD at 550 nm, whereas AERONET AOD does not have such a wavelength. Consequently, band interpolation is required. The AERONET AODs at 500 and 675 nm are interpolated into AOD at 550 nm using the Ångström exponent.

\[ \tau_{550} = \tau_{675} (550/675)^{-\alpha} \]  

\[ \alpha = -\ln(\tau_{500}/\tau_{675}) / \ln(500/675) \]  

where \( \tau_{500} \) denotes the AOD at 500 nm, \( \tau_{675} \) represents the AOD at 675 nm, \( \alpha \) is the Ångström exponent at 500–675 nm, and \( \tau_{550} \) indicates the AOD at 550 nm obtained by interpolation.

Considering that Sentinel-2 L1C products are resampled at different ground sampling distances (10, 20, and 60 m/pixel) based on the native resolution of different spectral bands [45], the L1C products must be first resampled to unify the spatial resolution before aerosol retrieval. In this study, all satellite images were resampled to 10 m/pixel. Satellite images are pixel-screened to remove samples containing light, clouds, cloud shadows, and snow from the dataset to improve the quality of samples and minimize the influence of the aforementioned factors. The AERONET AOD data after band interpolation are selected to match with the screened satellite images. Assuming that the aerosol concentration does not change in a short period of time, the average sun-photometer AOD value within ±30 min of the satellite overpass is selected to match with the satellite image. Assuming that the aerosol properties do not change within an area of 1.28 km × 1.28 km, the Sentinel-2 image within the range of 128 × 128 pixels is cropped as a satellite image sample with the AERONET site in the center. The satellite image and ground-based AOD data in the same spatiotemporal are paired together to provide samples for the model.

When training deep learning models, data augmentation methods, such as rotation, scaling, and translation, are frequently used to produce more samples to improve the generalization ability and robustness of the model [46]. The remote sensing images are processed by a horizontal translation of 0.1 to expand the dataset, enabling the model to better learn features from these images and avoid model overfitting. In this study, 2575 pairs of samples were successfully matched from the AERONET AOD data and Sentinel-2 images. Before training, all image samples are resized to 128 × 128 pixels. A total of 2575 pairs of samples are randomly assigned to the training dataset (2060 pairs) and the test dataset (515 pairs) in an 8:2 ratio.
samples are randomly assigned to the training dataset (2060 pairs) and the test dataset (515 pairs) in an 8:2 ratio.

Figure 2. Flowchart of the CNN model for the Sentinel-2 AOD retrieval.

3.2. CNN Model

CNN, which is a deep learning algorithm, consists of a series of convolution and pooling layers that extract relevant features from the input image, followed by one or more fully connected layers that use these features to make a prediction. It has the characteristics of sparse connections, parameter sharing, and translation invariance, greatly minimizing the number of network parameters that need to be optimized, improving the training speed, and facilitating the extraction of partial features in images. CNN is able to capture spatial features and patterns in images using a hierarchical architecture of layers that perform convolution operations and extract features at different levels of abstraction [47]. So, this algorithm has remarkable advantages for image analysis and recognition. The CNN model proposed in this work consists of 10 convolution layers, 4 max-pooling layers, 1 Global max-pooling layer, and 1 fully connected layer, as shown in Figure 3. Every convolutional block is followed by a max-pooling layer to minimize the number of parameters and compress image features to reduce computation and memory consumption. The last max-pooling layer is followed by a Global max-pooling layer, which can flatten the multidimensional image to obtain a feature vector with a size of 512. In addition, dropout regularization is used to randomly turn off some neurons with a certain probability during each iteration of training of the neural network [48]. Accordingly, each neuron does not necessarily participate in network training every time, and the updating of network weights no longer depends on the joint action of implicit nodes with fixed relationships. This approach prevents some features from being effective only under other specific features, reduces the dependence of model on some partial features, and enhances the generalization ability of the model. All convolutional layers use small-size $3 \times 3$ convolutional filters with size
1 stride and select the “same” padding to maintain the spatial resolution of the feature map. All max-pooling operations are performed over a window of $2 \times 2$ pixels with size 2 stride and the “same” padding to minimize the spatial resolution of the feature map by half.

![Figure 3. The network architecture of Sentinel-2 AOD retrieval.](image)

The loss function, activation function, and optimizer should be configured before training. After the training data of every iteration is inputted into the model, the model will output the predicted value through forward propagation, and the loss function will calculate the difference between the predicted and the true values. After the loss value is obtained, the model will update each parameter through back propagation to minimize the loss. Accordingly, the predicted value output by the model will be closer to the true value. Mean Absolute Error (MAE) was used as the loss function of the model in this study. The nonlinear activation function ReLU is selected as the activation function to minimize the interdependence between parameters. During the back propagation, the optimizer can guide the model parameters to update in the correct direction. Consequently, the updated parameters maintain the loss value close to the global minimum. The Adam optimization algorithm is used for network training, and the learning rate is set to 0.001 to minimize the MAE between retrieved AOD and true AOD. After the CNN is trained, the model with the smallest MAE between CNN AOD and AERONET AOD is selected as the final CNN model.

### 3.3. Evaluation Methods

Spearman R, Pearson R, $R^2$, Root Mean Square Error (RMSE), MAE, Within the MSI Expected Error (Within EE), Above EE, and Below EE were used as evaluation indicators in this study. Pearson R is sensitive to outliers and suitable for evaluating linear relationships. On the other hand, Spearman R is not affected by outliers and is suitable for evaluating nonlinear relationships. The model performance can be effectively evaluated using these two correlation coefficients as evaluation indexes.

\[
\text{Spearman R} = \frac{\sum (\text{AOD}_{\text{true}} - \overline{\text{AOD}}_{\text{true}}) (\text{AOD}_{\text{retrieved}} - \overline{\text{AOD}}_{\text{retrieved}})}{\sqrt{\sum \text{AOD}_{\text{true}}^2 - \overline{\text{AOD}}_{\text{true}}^2} \sqrt{\sum \text{AOD}_{\text{retrieved}}^2 - \overline{\text{AOD}}_{\text{retrieved}}^2}}
\]

\[
\text{Pearson R} = \frac{N \sum \text{AOD}_{\text{true}} \cdot \text{AOD}_{\text{retrieved}} - \sum \text{AOD}_{\text{true}} \sum \text{AOD}_{\text{retrieved}}}{\sqrt{N \sum \text{AOD}_{\text{true}}^2 - \sum \text{AOD}_{\text{true}}^2} \sqrt{N \sum \text{AOD}_{\text{retrieved}}^2 - \sum \text{AOD}_{\text{retrieved}}^2}}
\]

\[
R^2 = 1 - \frac{\sum (\text{AOD}_{\text{true}} - \text{AOD}_{\text{retrieved}})^2}{\sum (\text{AOD}_{\text{true}} - \overline{\text{AOD}}_{\text{true}})^2}
\]
Random Forest (RF) and VGG16 algorithms are selected for comparison and verification with the CNN model to verify the rationality of the CNN model. RF is an ensemble learning algorithm [49] that can evaluate the importance of every feature in classification problems, judge the interaction between different features, and capture the complex relationship between input feature vectors and output values. VGG is a classic CNN structure that is widely used in computer vision tasks, such as image classification, object detection, and semantic segmentation. VGG16 is a network structure in VGG [50].

4. Results and Discussion

4.1. Overall Validation of the CNN Model

AERONET AODs were used to compare and analyze the AOD retrieved by the CNN model on the training and test datasets. In the training dataset, the retrieved AOD is in good agreement with AERONET AOD, with an $R^2$ of 0.99, RMSE of 0.021, and Within EE of 99%, as shown in Figure 4. Only 1% of the data is higher than EE, indicating a very slight overestimation of the model in the training dataset. The result of CNN performance on the test dataset is similar to that on the training dataset. In the test dataset, $R^2$ is 0.95, RMSE is 0.049, and Within EE is 95%, indicating that CNN can provide accurate AOD concentration and distribution at a relatively fine temporal resolution (5 days). Approximately 3% of AODs are higher than EE, and 2% of AODs are lower than EE, indicating a slight deviation in the CNN AOD.

In the satellite images, the top-of-atmosphere signals are composed of coupled contributions from the surface and atmosphere, hence, the signal-to-noise-ratio of the high-AOD aerosols is larger than that of the low-AOD aerosols [35]. In the test dataset, 52% of the AERONET AODs are in the range of 0 to 0.16, and 48% are in the range of 0.16 to 1.6. The test dataset was divided into two parts to examine the accuracy of the CNN model in

\[
RMSE = \sqrt{\frac{1}{N} \sum (\text{AOD}_{\text{True}} - \text{AOD}_{\text{Retrieval}})^2}
\]  
(6)

\[
\text{MAE} = \frac{1}{N} \sum |\text{AOD}_{\text{True}} - \text{AOD}_{\text{Retrieval}}|
\]  
(7)

\[
\text{EE} = \pm (0.15 \times \text{AOD}_{\text{True}} + 0.05)
\]  
(8)

Figure 4. Scatterplots of the CNN AOD against AERONET AOD for the (a) training dataset and (b) test dataset. The dot color signifies the point density from blue (low) to red (high). The dashed black lines are the 1:1 line, and the dashed red lines are the EE envelopes $\pm (0.05 + 15\%)$. 

In the satellite images, the top-of-atmosphere signals are composed of coupled contributions from the surface and atmosphere, hence, the signal-to-noise-ratio of the high-AOD aerosols is larger than that of the low-AOD aerosols [35]. In the test dataset, 52% of the AERONET AODs are in the range of 0 to 0.16, and 48% are in the range of 0.16 to 1.6. The test dataset was divided into two parts to examine the accuracy of the CNN model in
retrieving low and high AODs: the low AOD test dataset ([0, 0.16]) and the high AOD test dataset ([0.16, 1.6]). Table 4 illustrates that MAE and RMSE in the low AOD are lower than those in the high AOD. On the other hand, the Within EE in the low AOD is similar to that in the high AOD. This finding indicates that CNN can better retrieve AOD in areas with good air quality. In the low AOD, 6% of AODs are higher than EE and no data is lower than EE. In the high AOD, 1% of AODs are higher than EE and 4% of AODs are lower than EE. The result shows that the CNN model slightly overestimates the low values and underestimates the high values.

Table 4. AOD retrieval performance of the CNN model for low and high AOD situations on the test dataset.

<table>
<thead>
<tr>
<th>Matched Number</th>
<th>MAE</th>
<th>RMSE</th>
<th>within EE</th>
<th>above EE</th>
<th>below EE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low AOD</td>
<td>269</td>
<td>0.021</td>
<td>0.034</td>
<td>94%</td>
<td>6%</td>
</tr>
<tr>
<td>High AOD</td>
<td>246</td>
<td>0.042</td>
<td>0.060</td>
<td>95%</td>
<td>1%</td>
</tr>
</tbody>
</table>

The AOD retrieval performance of the CNN model on the test dataset was subdivided by season and land cover type. CNN performs better in summer and winter than in spring and autumn (Table 5). From the perspective of the various validation indicators, the performance of CNN in winter is the best, with an $R^2$, MAE, RMSE, and Within EE of 0.97, 0.024, 0.037, and 97%, respectively. The excellent CNN performance in winter may be related to the large amount of winter data. The performance of CNN in summer is the second best, with an $R^2$ of 0.96 and RMSE of 0.045. On the other hand, the performance of CNN in autumn is similar to that in summer, with an $R^2$ of 0.94 and RMSE of 0.046. The CNN performance in spring is poorer compared with those in the other three seasons, with an $R^2$ of 0.93 and RMSE of 0.062.

Table 5. CNN AOD retrieval errors on the test dataset with respect to seasons.

<table>
<thead>
<tr>
<th>Seasons</th>
<th>Matched Number</th>
<th>Spearman R</th>
<th>Pearson R</th>
<th>$R^2$</th>
<th>MAE</th>
<th>RMSE</th>
<th>within EE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>144</td>
<td>0.95</td>
<td>0.97</td>
<td>0.93</td>
<td>0.039</td>
<td>0.062</td>
<td>94%</td>
</tr>
<tr>
<td>Summer</td>
<td>88</td>
<td>0.92</td>
<td>0.98</td>
<td>0.96</td>
<td>0.034</td>
<td>0.045</td>
<td>93%</td>
</tr>
<tr>
<td>Autumn</td>
<td>112</td>
<td>0.9</td>
<td>0.97</td>
<td>0.94</td>
<td>0.030</td>
<td>0.046</td>
<td>93%</td>
</tr>
<tr>
<td>Winter</td>
<td>171</td>
<td>0.95</td>
<td>0.99</td>
<td>0.97</td>
<td>0.024</td>
<td>0.037</td>
<td>97%</td>
</tr>
</tbody>
</table>

CNN efficiently performs AOD retrieval on the different land cover types (Table 6). CNN exhibits excellent retrieval accuracy in the vegetation surface, with an $R^2$ of 0.92 and RMSE of 0.044. When the surface reflectance is high, the satellite sensors can obtain less information about aerosols, making it difficult to retrieve aerosols on high-reflectance surfaces [51]. However, CNN still shows great retrieval performance on land cover types with high surface reflectance, such as urban and bare soil, with an $R^2$ of 0.96 and 0.90 and RMSE of 0.051 and 0.042, respectively.

Table 6. CNN AOD retrieval errors on the test dataset with respect to land cover types.

<table>
<thead>
<tr>
<th>Land Cover Types</th>
<th>Matched Number</th>
<th>Spearman R</th>
<th>Pearson R</th>
<th>$R^2$</th>
<th>MAE</th>
<th>RMSE</th>
<th>within EE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetation surface</td>
<td>128</td>
<td>0.94</td>
<td>0.96</td>
<td>0.92</td>
<td>0.029</td>
<td>0.044</td>
<td>95%</td>
</tr>
<tr>
<td>Urban</td>
<td>331</td>
<td>0.94</td>
<td>0.98</td>
<td>0.96</td>
<td>0.032</td>
<td>0.051</td>
<td>95%</td>
</tr>
<tr>
<td>Bare soil</td>
<td>56</td>
<td>0.9</td>
<td>0.95</td>
<td>0.90</td>
<td>0.030</td>
<td>0.042</td>
<td>95%</td>
</tr>
</tbody>
</table>

4.2. Retrieval Performance of the CNN Model at Different Scales

The AOD retrieval performance of the CNN model on the test dataset was subdivided into two scale types: regional scale and ground-based AERONET site scale, to analyze CNN performance in different regions and sites. The retrieval performance of CNN in North
China, Taiwan, South Korea, and Japan is shown in Figure 5. There is no analysis of Tibet and Hong Kong here because of the small amount of data in these regions. CNN performed best in South Korea, with an RMSE of 0.037, $R^2$ of 0.92, and Within EE of 98%. Moreover, 2% of AODs are lower than EE, indicating a slight underestimation of CNN in South Korea. In North China, CNN could accurately retrieve AOD with an $R^2$ of 0.97, RMSE of 0.047, and Within EE of 96%. Furthermore, 2% of AODs are higher than EE, whereas 2% are lower than EE, indicating a slight deviation of CNN AOD in North China. The more data in the region, the more regional features the model could capture. The largest amount of data can be found in North China. However, the retrieval performance of CNN in North China is not as good as that in South Korea, which may be because the aerosol concentration in North China is complex and changeable, and it is difficult for CNN to learn the aerosol characteristics in this region. CNN did not perform well in Taiwan, with its RMSE being the highest of the four regions. CNN performed the worst in Japan, with an $R^2$ of 0.78, RMSE of 0.045, and Within EE of 91%, which may be related to the smaller amount of data obtained in Japan. The annual average AOD in Japan is the lowest, and the CNN model slightly overestimates the low AODs with an Above EE of 6% (Table 4), so the CNN tends to overestimate the AOD in Japan, which has an Above EE of 8% and is significantly higher than those in North China and South Korea.

Figure 5. Scatterplots of the CNN AOD against AERONET AOD for the test dataset at different regional scales. The dot color signifies the point density from blue (low) to red (high). The dashed black lines are the 1:1 line, and the dashed red lines are the EE envelopes $\pm (0.05 + 15\%)$. 

Spearman R = 0.94 Pearson R = 0.99
MAE = 0.031 RMSE = 0.047
Within EE = 96%
Above EE = 2%
Below EE = 2%
N = 234

North China

Spearman R = 0.94 Pearson R = 0.97
MAE = 0.041 RMSE = 0.066
Within EE = 90%
Above EE = 7%
Below EE = 2%
N = 81

Taiwan region

Spearman R = 0.95 Pearson R = 0.96
MAE = 0.025 RMSE = 0.037
Within EE = 98%
Above EE = 0%
Below EE = 2%
N = 84

South Korea

Spearman R = 0.85 Pearson R = 0.88
$R^2$ = 0.78
MAE = 0.029 RMSE = 0.045
Within EE = 91%
Above EE = 8%
Below EE = 1%
N = 77

Japan
The CNN performance at each site is shown in Figure 6. Given the limited amount of data at TB01 and KR1, only one TB01 data and two KR1 data were divided into the test dataset. Consequently, the retrieval performance at these two sites is not shown. The AOD retrieval accuracy at NC02 is the highest, with an $R^2$, RMSE, and Within EE of 0.98, 0.027, and 99%, respectively. This result may be due to the largest amount of data being at NC02, with up to 311 pairs of samples, from which CNN could capture more accurate image features. Moreover, the land cover type at NC02 is simple and mainly urban. In addition, CNN performed well at TW01, with its RMSE being the lowest among all sites (0.016). The land cover type at TW01 is simple, with the vast majority of the area covered by vegetation, and CNN can obtain more aerosol information in an area with low surface reflectance. The AOD retrieval accuracy at JP4 is the lowest, with an $R^2$ of 0.73, RMSE of 0.095, and Within EE of 67%. CNN overestimated the AOD at JP4 with an Above EE of 33%. The low accuracy of CNN AOD at JP4 is due to two reasons. First, high AOD more frequently occurs in China and South Korea than in Japan [52], and the annual average AOD in Japan is the lowest, so CNN tends to overestimate the AOD in Japan. Second, the data at JP4 are few (only 52 pairs of samples). The deep learning model requires a large number of samples, and CNN would have difficulty exploring the characteristics of the site using only a few samples. In addition, CNN did not perform well at TW03, with its RMSE being the highest of all sites (0.122). TW03 is characterized by a relatively complex land cover type, primarily dominated by urban areas with some vegetation coverage. Moreover, the amount of data obtained at TW03 is small, with only 55 pairs of samples. Accordingly, the accuracy of CNN AOD at TW03 is low. CNN exhibits a slight AOD deviation at most sites. Nonetheless, the CNN AOD at HK01, HK02, KR4, and JP3 reaches 100% Within EE, and no AOD is higher or lower than EE at all. CNN retrieved AOD better at sites with more data, indicating that CNN can obtain more features at sites with more data. Expanding the dataset may help the model to better extract image features related to aerosols, and a large dataset could be made for model training in the future.

4.3. Comparison with Other Models

The MODIS and Himawari-8 aerosol products were selected to evaluate the rationality of the spatial distribution of CNN AOD. Figure 7a shows a Sentinel-2 satellite image over Chiba and Ibaraki in Japan on 12 February 2019, with no snow and clouds in most areas, making it suitable for AOD retrieval. Figure 7c–g shows MODIS MAIAC, DT, DB, DTB, and Himawari-8 aerosol products over Japan, with spatial resolutions of 1 km/pixel, 10 km/pixel, 10 km/pixel, 10 km/pixel, and 0.05 $°$/pixel, respectively. Figure 7b shows the spatial distribution of AOD retrieved by CNN. CNN AODs are calculated with a step size of 64 pixels and an input image size of 128 $\times$ 128 to facilitate display and comparison, resulting in a spatial resolution of 640 m/pixel. Figure 7a depicts some cloud and snow cover in the lower left corner and middle upper areas of the image. Incorrect AODs are retrieved by CNN due to the interference of clouds and snow. Except for the cloud and snow coverage area, the spatial distribution of CNN AOD and MODIS AOD is similar, and the overall AOD concentration of the image coverage area is low. Some coastal areas show relatively high AODs due to dense human habitation and intensive industrial agglomeration. On the other hand, inland and southern mountain areas show relatively low AODs. The Himawari-8 aerosol product shows an overall higher AOD, which is not consistent with MODIS AOD and CNN AOD. The MODIS AOD product is more reliable than the Himawari-8 AOD product [34]. Specifically, the CNN AOD that is more similar to MODIS AOD can reasonably reflect the AOD concentration. The distribution maps of the CNN AOD and Himawari-8 AOD cover the entire land. The data coverage rate of the MODIS aerosol products is low, and many no-value regions exist. The CNN AOD and Himawari-8 AOD can provide continuous AOD spatial distribution. Furthermore, the high spatial resolution map of the CNN AOD shows the spatial details that cannot be reflected in the rough spatial resolution of 1 km/pixel, 10 km/pixel, and 0.05 $°$/pixel. The AOD retrieved by CNN is closer to the ground-based AOD, with an RMSE of 0.049. In contrast,
the RMSE values of MODIS MAIAC, DT, DB, and Himawari-8 AOD are 0.15, 0.22, 0.17, and 0.14, respectively, which are less accurate than CNN.

Figure 6. Scatterplots of the CNN AOD against AERONET AOD for test dataset at each site. The dashed black lines are the 1:1 line, and the dashed red lines are the EE envelopes ± (0.05 + 15%).
The CNN model was compared with RF and VGG16 to further evaluate the AOD retrieval performance of CNN. The RF and VGG16 algorithms were trained and tested using the dataset produced in this study. The output layer of VGG16 was changed to a fully connected layer with size 1 because it is a classification model that eventually divides into 1000 categories. The AOD retrieval performance of each model is shown in Table 7. The results of MAIAC, DT, DB, and YAER are from a study by Choi et al. [53], in which the MODIS and Himawari-8 aerosol products in East Asia are validated. CNN shows outstanding advantages in AOD retrieval, and all verification indicators are the best values, with an RMSE of 0.049, and significantly lower than those of other models. The $R^2$ of CNN is 0.95, which is higher than those of other models. RF performs well, with an $R^2$ of 0.86, which is higher than the 0.70 of VGG16. The RMSE of RF is 0.085, which is lower than the 0.123 of VGG16. The results show that CNN demonstrates significant advancements in addressing AOD retrieval problems compared with most previous algorithms.

Table 7. AOD retrieval performance of aerosol products and models.

<table>
<thead>
<tr>
<th>Method</th>
<th>Satellite</th>
<th>Spearman R</th>
<th>Pearson R</th>
<th>$R^2$</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>Sentinel-2</td>
<td>0.94</td>
<td>0.98</td>
<td>0.95</td>
<td>0.031</td>
<td>0.049</td>
</tr>
<tr>
<td>MAIAC</td>
<td>Terra</td>
<td>/</td>
<td>0.93</td>
<td>/</td>
<td>/</td>
<td>0.150</td>
</tr>
<tr>
<td>DT</td>
<td>Terra</td>
<td>/</td>
<td>0.87</td>
<td>/</td>
<td>/</td>
<td>0.220</td>
</tr>
<tr>
<td>DB</td>
<td>Terra</td>
<td>/</td>
<td>0.88</td>
<td>/</td>
<td>/</td>
<td>0.170</td>
</tr>
<tr>
<td>YAER</td>
<td>Himawari-8</td>
<td>/</td>
<td>0.91</td>
<td>/</td>
<td>/</td>
<td>0.140</td>
</tr>
<tr>
<td>VGG16</td>
<td>Sentinel-2</td>
<td>0.79</td>
<td>0.86</td>
<td>0.70</td>
<td>0.079</td>
<td>0.123</td>
</tr>
<tr>
<td>RF</td>
<td>Sentinel-2</td>
<td>0.88</td>
<td>0.93</td>
<td>0.86</td>
<td>0.054</td>
<td>0.085</td>
</tr>
</tbody>
</table>

$//_*$ means not given in the literature.

4.4. Retrieval Performance as a Function of Satellite Image Size

Image datasets of $32 \times 32$, $64 \times 64$, and $128 \times 128$ pixels were constructed to explore the relationship between image size and AOD retrieval performance. The model was retrained with datasets of different sizes, and the hyperparameters of CNN were adjusted. The smaller the size, the less information the satellite image contains. Accordingly, the
image feature vector of the fully connected layer in the CNN was reduced to 80 when the model was trained with datasets of 64 and 32 sizes. The AOD retrieval performance on the dataset with different sizes is shown in Table 8. The 128-size model works best, with an $R^2$ of 0.95, which is 0.03 and 0.02 higher than those of the 32-size model and 64-size model, respectively. The RMSE of the 128-size model is 0.049, which is 0.016 and 0.011 lower than those of the 32-size model and 64-size model, respectively. The 32-size model performed worst, with the lowest $R^2$, highest MAE, and highest RMSE among the three models, and its overestimation and underestimation are the most severe. The better results of the model trained with a larger image dataset indicate that the 128-size image contains richer aerosol information.

Table 8. CNN AOD-retrieval errors on the test dataset with sizes of 32, 64, and 128.

<table>
<thead>
<tr>
<th>Image Size</th>
<th>Spearman R</th>
<th>Pearson R</th>
<th>$R^2$</th>
<th>MAE</th>
<th>RMSE</th>
<th>within EE</th>
<th>above EE</th>
<th>below EE</th>
</tr>
</thead>
<tbody>
<tr>
<td>32 $\times$ 32</td>
<td>0.94</td>
<td>0.96</td>
<td>0.92</td>
<td>0.039</td>
<td>0.065</td>
<td>91%</td>
<td>5%</td>
<td>4%</td>
</tr>
<tr>
<td>64 $\times$ 64</td>
<td>0.95</td>
<td>0.97</td>
<td>0.93</td>
<td>0.032</td>
<td>0.060</td>
<td>94%</td>
<td>4%</td>
<td>2%</td>
</tr>
<tr>
<td>128 $\times$ 128</td>
<td>0.94</td>
<td>0.98</td>
<td>0.95</td>
<td>0.031</td>
<td>0.049</td>
<td>95%</td>
<td>3%</td>
<td>2%</td>
</tr>
</tbody>
</table>

5. Conclusions

In this study, a novel AOD-retrieval model based on CNN was developed, which can directly retrieve AOD from Sentinel-2 satellite images, improving the stability and spatiotemporal adaptability of aerosol retrieval. The CNN model consists of 10 convolutional layers, 4 max-pooling layers, 1 Global max-pooling layer, and 1 fully connected layer. The image features related to aerosol load in Sentinel-2 images of four bands were extracted using the excellent image feature extraction function of CNN. The correlation between the AOD and the remote sensing images was modeled to achieve AOD retrieval. Taking East Asia as the research area, the data of 22 AERONET sites from 2016 to 2022 were obtained, and the satellite image and ground-based AOD data in the same spatiotemporal were paired together to make 2575 pairs of samples. The samples were enhanced by a horizontal shift of 0.1. The model was trained and tested by selecting data in a ratio of 8:2 from samples. The results are as follows:

(1) The proposed model can accurately retrieve AOD, with an $R^2$ of 0.95, RMSE of 0.049, and Within EE of 95% on the test dataset. The AOD-retrieval accuracy of CNN is higher compared with those of the DT, DB, DB, MAIAC, YAER, RF, and VGG16 algorithms. In addition, CNN could provide continuous and detailed aerosol distribution to fill the observation gap in existing ground-based monitoring networks.

(2) CNN efficiently performs AOD retrieval on different land cover types: vegetation surface, urban, and bare soil. When the surface reflectance is high, the satellite sensors can obtain less information about aerosols, making it difficult to retrieve aerosols on a high-reflectance surface. However, CNN still shows great AOD retrieval potential on surfaces with high surface reflectance, such as urban and bare soil, with an $R^2$ of 0.96 and 0.90 and RMSE of 0.051 and 0.042, respectively.

(3) CNN performs better in summer and winter than in spring and autumn. The performance of CNN in winter is the best, with an $R^2$ of 0.97 and RMSE of 0.037. The performance of CNN in summer is the second best. The CNN performance in spring is poorer compared with those in the other three seasons, with an $R^2$ of 0.93 and RMSE of 0.062.

(4) To investigate the relationship between image size and model retrieval performance, datasets of 32, 64, and 128 sizes were created to train and test the CNN. The 128-size CNN performed better because of the rich AOD information in the 128-size image.
Author Contributions: Conceptualization, methodology, validation, writing, visualization, J.J. and J.L.; writing—review and editing, D.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NUPTSF, grant number NY20211.


Acknowledgments: The authors thank the Pls, site managers, and other staff for operations in AERONET sites used in this investigation. They also thank the Finance science and technology project of Hainan province (ZDYF2020206) for its support with AERONET work. The authors are also grateful to ESA for providing Sentinel-2 data. The authors acknowledge MODIS and Himawari-8 groups for the aerosol products.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Xue, Y.; He, X.W.; Xu, H.; Guang, J.P.; Guo, J.P.; Mei, L.L. China Collection 2.0: The aerosol optical depth dataset from the synergetic retrieval of aerosol properties algorithm. Atmos. Environ. 2014, 95, 45–58. [CrossRef]
9. Twomey, S. Pollution and the planetary albedo. Atmos. Environ. 1974, 8, 1251–1256. [CrossRef]
14. Matkovic, V.; Mulić, M.; Azabagić, S.; Jevtić, M. Premature Adult Mortality and Years of Life Lost Attributed to Long-Term Exposure to Ambient Particulate Matter Pollution and Potential for Mitigating Adverse Health Effects in Tuzla and Lukavac, Bosnia and Herzegovina. Atmosphere 2020, 11, 1107. [CrossRef]


36. Wei, J.; Yang, F.; Ren, X.C.; Zou, S. A Short-Term Prediction Model of PM2.5 Concentration Based on Deep Learning and Mode Decomposition Methods. *Appl. Sci.* 2021, 11, 6915. [CrossRef]

37. Yuan, H.; Xu, G.; Lv, T.; Ao, A.; Zhang, Y. PM2.5 Forecast Based on a Multiple Attention Long Short-Term Memory (MAT-LSTM) Neural Networks. *Anal. Lett.* 2020, 54, 935–946. [CrossRef]

38. Yan, X.; Zhang, Z.; Luo, N.; Jiang, Y.; Li, Z. New interpretable deep learning model to monitor real-time PM2.5 concentrations from satellite data. *Environ. Int.* 2020, 144, 106060. [CrossRef] [PubMed]


42. Li, G.; Mao, Y.; Feng, X.; Chen, Z.; Yang, Z.; Cheng, X. Monitoring ice flow velocity of Petermann glacier combined with Sentinel-1 and –2 imagery. *Int. J. Appl. Earth Obs.* 2023, 121, 1569–8432. [CrossRef]


45. Yang, Y.; Chen, Y.; Yang, K.; Cermak, J.; Chen, Y. High-resolution aerosol retrieval over urban areas using sentinel-2 data. *Atmos. Res.* 2021, 264, 105829. [CrossRef]
47. Krichen, M. Convolutional Neural Networks: A Survey. *Computers* 2023, 12, 151. [CrossRef]
52. Kang, Y.; Kim, M.; Kang, E.; Cho, D.; Im, J. Improved retrievals of aerosol optical depth and fine mode fraction from GOCI geostationary satellite data using machine learning over East Asia. *ISPRS J. Photogramm.* 2022, 183, 253–268. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.