Scientific
Research
Publishing

Effect of the Pixel Interpolation Method for Downsampling Medical Images on Deep Learning Accuracy

Daisuke Hirahara[1,2,3], Eichi Takaya[4], Mizuki Kadowaki[2], Yasuyuki Kobayashi[3], Takuya Ueda[2]

[1]Department of AI Research Lab, Harada Academy, Kagoshima, Japan

[2]Department of Clinical Imaging, Graduate School of Medicine, Tohoku University, Sendai, Japan

[3]Department of Advanced Biomedical Imaging Informatics, St. Marianna University School of Medicine, Kawasaki, Japan

[4]Tohoku University Hospital AI Lab, Sendai, Japan

Email: ffieldai@gmail.com

Abstract

Background: High-resolution medical images often need to be downsampled because of the memory limitations of the hardware used for machine learning. Although various image interpolation methods are applicable to downsampling, the effect of data preprocessing on the learning performance of convolutional neural networks (CNNs) has not been fully investigated. **Methods:** In this study, five different pixel interpolation algorithms (nearest neighbor, bilinear, Hamming window, bicubic, and Lanczos interpolation) were used for image downsampling to investigate their effects on the prediction accuracy of a CNN. Chest X-ray images from the NIH public dataset were examined by downsampling 10 patterns. **Results:** The accuracy improved with a decreasing image size, and the best accuracy was achieved at 64×64 pixels. Among the interpolation methods, bicubic interpolation obtained the highest accuracy, followed by the Hamming window.

1. Introduction

Convolutional neural networks (CNNs) are a deep learning approach, and developments have increasingly widened their range of practical application [1].

CNNs are widely used for medical imaging applications, such as detecting lung tumors in computed tomography images, detecting breast cancer in mammograms, and predicting the risk for cardiovascular disease based on retinal fundus photographs [2] [3] [4]. Their adoption in the clinical field makes ensuring their accuracy imperative. For a CNN to yield correct diagnoses, they require image data with sufficient spatial resolution to extract morphological features [5]. Because of practical limitations, medical images often do not provide sufficient spatial resolution [6]. Certain examination processes such as magnetic resonance imaging may require a patient to hold their breath, which limits the imaging duration. Target lesions often occupy a localized or limited part of an organ or structure in the human body, so the region of interest is much smaller than the overall image, and the cropped image does not have sufficient spatial resolution [7]. When images with insufficient spatial resolution are used as training data, the CNN does not achieve satisfactory diagnostic accuracy. Some studies have shown that using pixel interpolation to upsample images is a useful solution for improving the diagnostic accuracy of CNNs when the training data have insufficient spatial resolution [8] [9].

In contrast, for some medical imaging applications such as pathological imaging and digital mammography, much higher resolutions compared to general images are sometimes used. The performance of a CNN generally improves when the batch size of the training data is increased [10]. However, researchers are often forced to limit the batch size because of the memory limitations of the hardware. Downsampling of high-resolution images is a possible solution to maintaining a larger batch size when the processing hardware has limited memory [11]. Sabottke *et al.* reported that training a CNN with chest radiographic images downsampled from 1024×1024 pixels to 256×256 pixels does not compromise the diagnostic accuracy [11]. Although various pixel interpolation algorithms are applicable to image downsampling such as nearest neighbor (NN) [12], bilinear (BL) [12], Hamming window (HM) [13], bicubic (BC) [14], and Lanczos (LC) [15] interpolation, the effect of the interpolation method on the diagnostic accuracy of a CNN has not been fully investigated.

In this study, five pixel interpolation algorithms (NN, BL, HM, BC, and LC) were applied to downsampling images, which were then used as training data for a CNN. The objective was to compare their effects on the diagnostic accuracy.

2. Materials & Methods

The medical images used to train the CNN were taken from the public dataset of NIH chest X-rays registered in Kaggle (Chest14) [16]. This dataset contains 112,120 images of 30,805 patients. Each chest X-ray is a grayscale image with a size of 1024×1024 pixels. The chest X-rays were classified according to 15 diagnostic labels: normal (no disease), atelectasis, consolidation, infiltration, pneumothorax, pneumothorax, edema, emphysema, fibrosis, effusion, pneumonia, pleural thickening, cardiomegaly, nodule mass, and hernia. Over 90% of the data contained an abnormal diagnosis [16]. The dataset used is open data and has

been anonymized.

PyTorch was used to build the CNN, and Pillow was used for the interpolation method with Python version 3.6.9. The hardware environment was a DGX Station (CPU, Intel Xeon E5-2698 v4 2.2 GHz; system memory of 256 GB, GPUs, Tesla V100 × 4; GPU memory, 32 GB per GPU).

The image data for the CNN were interpolated by using five pixel interpolation methods in the image-processing library of Python Pillow: NN [12], BL [12], HM [13], BC [14], and LC [15]. NN is a pixel interpolation algorithm that refers to the brightness value of the pixel nearest to the reference position [12]. BL is a linear interpolation algorithm that resamples the luminance value of a position (x, y) by referring to luminance values for 2 × 2 pixels (4 pixels) in the surrounding vicinity [12]. HN and LC are commonly used window functions. HM is a modified version of the Hanning window, where the values at both ends of the window are zero, so the signal components are not reflected in the spectrum [13]. BC is a cubic interpolation algorithm that resamples the luminance value of a position by referring to luminance values for a bicubic array of 4 × 4 pixels (16 pixels) [14]. HM and LC offer a better frequency resolution and narrower dynamic range than the Hanning window. LC is characterized by discontinuities at both ends of the interval and is one of many finite support approximations of the sinc filter. Each interpolated value is the weighted sum of two consecutive input samples [15].

The five interpolation methods were applied to downsampling the chest X-rays from an original size of 1024 × 1024 pixels to seven different sizes: 320 × 320, 256 × 256, 224 × 224, 192 × 192, 160 × 160, 64 × 64, and 32 × 32 pixels. In total, 112,120 images were generated and were separated into 86,524 training images and 25,596 testing images. Verification was not performed with the original data because the batch size could not be normalized with the other batch sizes owing to the amount of spatial computation. The sizes of the images used as input for the CNN were 320 × 320, 256 × 256, 224 × 224, 192 × 192, 160 × 160, 64 × 64, and 32 × 32 pixels. As shown in Figure 1, the CNN was a simple

$$O_h = \frac{H + 2P - Fh}{S} + 1$$
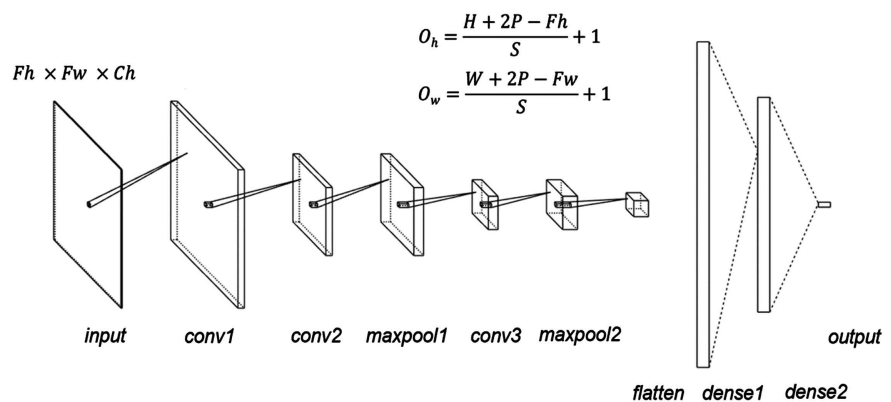
$$O_w = \frac{W + 2P - Fw}{S} + 1$$

Figure 1. CNN structure (Fh: input height, Fw: input width, Oh: output height, Ow: output width, P: padding, S: stride; kernel size: 5, stride: 1, padding: 0, dropout: 0.5).

network consisting of three 3 × 3 convolutional layers. It used zero padding, instance normalization, and a rectified linear unit activation function. To accommodate the above image sizes of the downsampled input data, the CNN was adjusted to have 360,000, 222,784, 166,464, 118,336, 78,400, 7744, or 576 fully connected layers respectively. For all resolutions, the batch size was fixed at 512. The CNN was trained on each image for 100 epochs in total, and the accuracy was evaluated after training and testing. The results were compared according to the maximum classification accuracy over 100 epochs.

3. Results

Table 1 and Figure 2 present the classification accuracy of the test data with each model. Reducing the image size increased the accuracy. The best accuracy was obtained at a size of 64 × 64 pixels and was slightly worse at a size of 32 × 32 pixels. At 320 × 320 pixels, BC performed the best with a maximum accuracy of
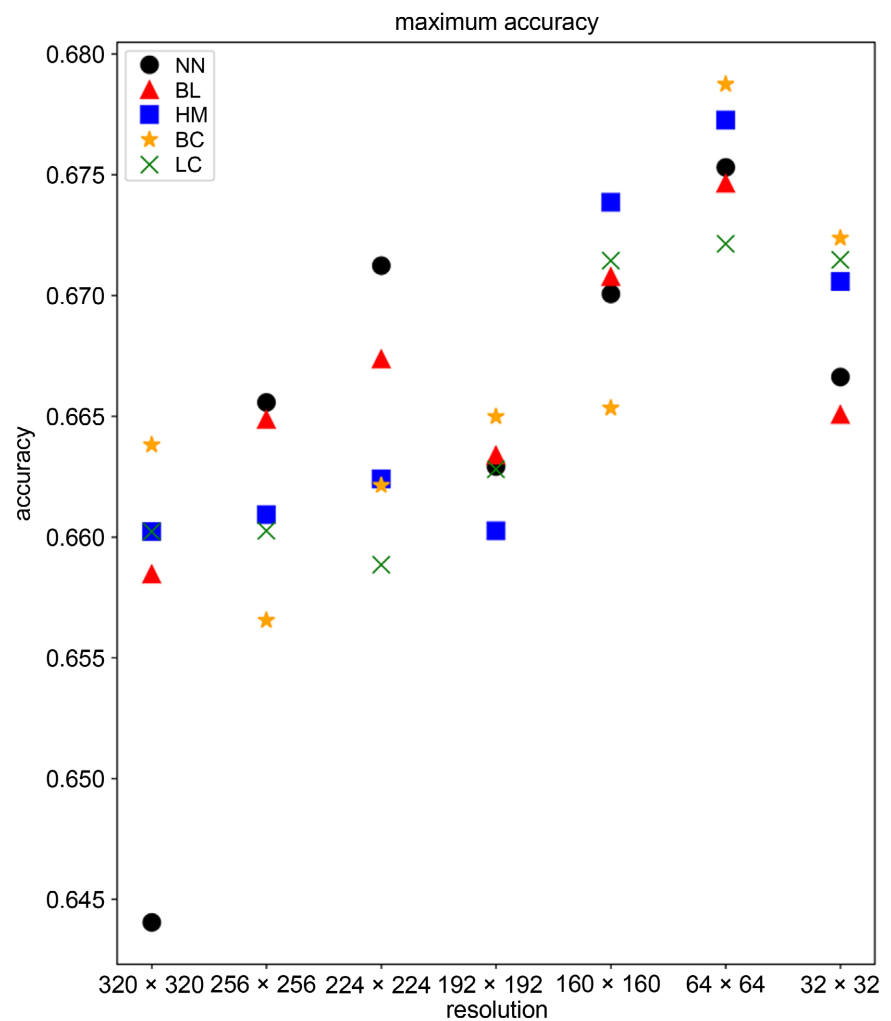

Figure 2. Maximum classification accuracy of the CNN according to the interpolation method and image size. (NN: Nearest neighbor, BL: Bilinear, HM: Hamming window, BC: Bicubic, LC: Lanczos).

Table 1. Maximum classification accuracy according to the interpolation method and image size.

	Max	Average
NN	0.6753	0.6686
BL	0.6746	0.6677
HM	0.6739	0.6675
BC	0.6787	0.6667
LC	0.6721	0.6662

0.6638. At 256 × 256 pixels, NN performed the best with a maximum accuracy of 0.6656. At 224 × 224 pixels, NN again performed the best with a maximum accuracy of 0.6712. At 192 × 192 pixels, BC performed the best with a maximum accuracy of 0.6650. At 160 × 160 pixels, HM performed the best with a maximum accuracy of 0.6739. At 64 × 64 pixels, BC performed the best with a maximum accuracy of 0.6787. At 32 × 32 pixels, BC performed the best with a maximum accuracy of 0.6724. Thus, the highest accuracy was obtained by BC at an image size of 64 × 64 pixels (0.6787), followed by HM at 64 × 64 pixels (0.6773).

4. Discussion

The results of this study suggest that the diagnostic performance of the CNN is affected by the level of downsampling. The maximum accuracy was achieved at an image size of 64 × 64 pixels regardless of the interpolation method. This supports the work of Tan *et al.* [17], who reported an optimal input dimension depending on the number of parameters for the deep learning model. The present study suggests that the optimal downsampling level is related to the CNN structure. The CNN used here had a relatively simple structure with three hidden layers, so it is potentially able to extract structural features with relatively low resolution. Because more complicated models are commonly used in clinical applications of artificial intelligence, the optimal downsampling process may need to be reassessed according to the CNN that is being used. The classification accuracy increased with a decreasing image size and reached its maximum at 64 × 64 pixels. Because the CNN had a simple structure with few hidden layers and little ability to extract features, downsampling to a small image size such as 64 × 64 or 32 × 32 pixels may serve the same function as feature extraction by the CNN. In other words, a small image size may avoid the curse of dimensionality.

The results also suggested that the interpolation method affects the diagnostic performance of the CNN differently according to the image size. Figure 2 partially reveals the relationship between the image size and interpolation method. For small image sizes such as 64 × 64 and 32 × 32 pixels, BC tended to produce good results. Linear interpolation methods such as NN and BL were not suitable for strong downsampling, while BC performed better, perhaps because it is nonlinear. Linear interpolation (NN, BL) may offer better performance when images are

downsampled to a certain size, and nonlinear interpolation (BC) may be more suitable when images are downsampled to extremely small sizes. Various interpolations are available for upsampling and downsampling images to different sizes. In the OpenCV and Pillow libraries, BL and BC are set as the default methods for image interpolation [https://pillow.readthedocs.io/en/stable/reference/Image.html]. BL was previously reported to result in the highest classification accuracy when images with a small size were upsampled [8]. In contrast, the present study suggests that BC results in the highest classification accuracy when images with a large size are downsampled. The two sets of results suggest that a suitable image interpolation method needs to be selected according to the image size and task (*i.e.*, upsampling of downsampling). The limitations of this study were that the effects of the model size and dataset were not considered. Future work will involve investigating larger network architectures and multiple types of medical image datasets. BC also performed better in large downsamplings, but the difference in classification accuracy between BC and BL at 64×64 pixels was only 0.41%. The difference between BC and NN is even smaller at 0.34%. More studies with different hyperparameters, such as changing the batch size for each resolution, are needed to clarify why this result is significant.

5. Conclusion

In this study, five different interpolation methods were applied to the Chest14 dataset, and the effect of image downsampling on the classification accuracy of the CNN was investigated. The experimental results showed that there is an optimal image size for downsampling, and that a suitable interpolation method should be chosen. The results of this study suggest the importance of considering the interpolation method when creating a deep learning model. It was shown that choosing hamming or bicubic for downsampling may give better accuracy than other interpolation methods.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Hijazi, S., Kumar, R. and Rowen, C. (2015) Using Convolutional Neural Networks for Image Recognition. Cadence Design Systems Inc., San Jose, 1-12.

[2] Lakwaa, W., Mohammad, N. and Amr, B. (2017) Lung Cancer Detection and Classification with 3D Convolutional Neural Network (3D-CNN). *Lung Cancer*, **8**, 409-417. https://doi.org/10.14569/IJACSA.2017.080853

[3] Sampaio, W.B., Diniz, E.M., Silva, A.C., de Paiva, A.C. and Gattass, M. (2011) Detection of Masses in Mammogram Images Using CNN, Geostatistic Functions and SVM. *Computers in Biology and Medicine*, **41**, 653-664. https://doi.org/10.1016/j.compbiomed.2011.05.017

[4] Poplin, R., Varadarajan, A.V., Blumer, K., Liu, Y., McConnell, M.V., Corrado, G.S.,

et al. (2018) Prediction of Cardiovascular Risk Factors from Retinal Fundus Photographs via Deep Learning. *Nature Biomedical Engineering*, **2**, 158-164.
https://doi.org/10.1038/s41551-018-0195-0

[5] Willemink, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., *et al.* (2020) Preparing Medical Imaging Data for Machine Learning. *Radiology*, **295**, 4-15. https://doi.org/10.1148/radiol.2020192224

[6] Powers, W.J., Rabinstein, A.A., Ackerson, T., Adeoye, O.M., Bambakidis, N.C., Becker, K., *et al.* (2018) 2018 Guidelines for the Early Management of Patients with Acute Ischemic Stroke: A Guideline for Healthcare Professionals from the American Heart Association/American Stroke Association. *Stroke*, **49**, e46-e99.
https://doi.org/10.1161/STR.0000000000000158

[7] Tomita, H., Yamashiro, T., Heianna, J., Nakasone, T., Kobayashi, T, Mishiro, S., *et al.* (2021) Deep Learning for the Preoperative Diagnosis of Metastatic Cervical Lymph Nodes on Contrast-Enhanced Computed ToMography in Patients with Oral Squamous Cell Carcinoma. *Cancers*, **13**, Article No. 600.
https://doi.org/10.3390/cancers13040600

[8] Hirahara, D., Takaya, E., Takahara, T. and Ueda, T. (2020) Effects of data count and image scaling on Deep Learning training. *PeerJ Computer Science*, **6**, Article No. e312. https://doi.org/10.7717/peerj-cs.312

[9] Alsallakh, B., Kokhlikyan, N., Miglani, V., Yua, J. and R-Richardson, O. (2020) Mind the Pad—CNNS Can Develop Blind Spots. arXiv preprint arXiv:2020.02178.

[10] Kandel, I. and Castelli, M. (2020) The Effect of Batch Size on the Generalizability of the Convolutional Neural Networks on a Histopathology Dataset. *ICT Express*, **6**, 312-315. https://doi.org/10.1016/j.icte.2020.04.010

[11] Sabottke, C.F. and Spieler, B.M. (2020) The Effect of Image Resolution on Deep Learning in Radiography. *Radiology: Artificial Intelligence*, **2**, Article ID: e190015.
https://doi.org/10.1148/ryai.2019190015

[12] Lehmann, T.M., Gönner, C. and Spitzer, K. (1999) Survey: Interpolation Methods in Medical Image Processing. *IEEE Transactions on Medical Imaging*, **18**, 1049-1075.
https://doi.org/10.1109/42.816070

[13] Harris, F.J. (1978) On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform. *Proceedings of the IEEE*, **66**, 51-83.
https://doi.org/10.1109/PROC.1978.10837

[14] Keys, R. (1981) Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **29**, 1153-1160.
https://doi.org/10.1109/TASSP.1981.1163711

[15] Duchon, C.E. (1979) Lanczos Filtering in One and Two Dimensions. *Journal of Applied Meteorology and Climatology*, **18**, 1016-1022.
https://doi.org/10.1175/1520-0450(1979)018%3C1016:LFIOAT%3E2.0.CO;2

[16] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. and Summers, R.M. (2017) ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 3462-3471. https://doi.org/10.1109/CVPR.2017.369

[17] Tan, M. and Quoc, V.L. (2019) EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv preprint arXiv:1905.11946.