International Journal of Geographical Information Science
Vol. 00, No. 00, Month 2021, 1–23

RESEARCH ARTICLE

Enriching the metadata of map images: a deep learning approach with GIS-based data augmentation

Yingjie Hu[1], Zhipeng Gui[2,3], Jimin Wang[1], and Muxian Li[2]
[1]GeoAI Lab, Department of Geography, University at Buffalo, SUNY
[2]School of Remote Sensing and Information Engineering, Wuhan University
[3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University

(*Received 00 Month 200x; final version received 00 Month 200x*)

Maps in the form of digital images are widely available in geoportals, Web pages, and other data sources. The metadata of map images, such as spatial extents and place names, are critical for their indexing and searching. However, many map images have either mismatched metadata or no metadata at all. Recent developments in deep learning offer new possibilities for enriching the metadata of map images via image-based information extraction. One major challenge of using deep learning models is that they often require large amounts of training data that have to be manually labeled. To address this challenge, this paper presents a deep learning approach with GIS-based data augmentation that can automatically generate labeled training map images from shapefiles using GIS operations. We utilize such an approach to enrich the metadata of map images by adding spatial extents and place names extracted from map images. We evaluate this GIS-based data augmentation approach by using it to train multiple deep learning models and testing them on two different datasets: a Web Map Service image dataset at the continental scale and an online map image dataset at the state scale. We then discuss the advantages and limitations of the proposed approach.

Keywords: Map image; map information extraction; geospatial metadata; data augmentation; GeoAI.

ISSN: 1365-8816 print/ISSN 1362-3087 online

DOI: 10.1080/1365881YYxxxxxxxx
http://www.informaworld.com

1. Introduction

A large number of map images are available in geoportals, historical archives, university libraries, Web pages, and other data sources. Some of these images were created by scanning paper-based maps. For example, the US Geological Survey (USGS) provides high-resolution scanned images of more than 200,000 historical topographic maps between 1884 and 2006 (Uhl *et al.* 2018). Some of these map images were created directly in digital format. For example, a GIS professional may use a mapping software, such as ArcGIS or QGIS, to create a map image. A geospatial data provider, such as NASA Socioeconomic Data and Applications Center (SEDAC), may render hosted geospatial data as map images and share them through online map services, such as the OGC Web Map Service (WMS) (Gui *et al.* 2013).

Given the large number of map images available, there exists a demand for indexing and searching them efficiently. Such a demand can be seen in geoportals, such as the U.S. GeoPlatform Portal and the INSPIRE geoportal in Europe, whose essential goal is to facilitate the discovery and reuse of the managed geospatial resources (Bernard *et al.* 2005, Lutz and Klien 2006). Meanwhile, many university libraries are transitioning to digital hubs, and also have the demand of enabling efficient search of their maintained materials and collections including map images (Wallis *et al.* 2010). In addition, a general Web search engine may want to return a set of relevant map images when a user types in a place name. This can already be seen on Google Image Search: when typing in a place name e.g., “California”, one can get many California maps among the returned results. However, these map images are limited to those that have the keyword “California” in its caption or in the content of the related Web page.

Metadata are of vital importance for efficiently searching and indexing map images in various geoportals and geospatial data repositories. Among the many elements in metadata, information about the geographic area covered by a map is essential for finding relevant map images. According to the Federal Geographic Data Committee (FGDC), metadata may contain Indirect Spatial Reference (ISR) and Direct Spatial Reference (DSR). ISR refers to the use of place names or addresses, which are often found in the textual descriptions of metadata (such as titles, keywords, and abstracts), to describe geographic areas. DSR refers to the use of geographic coordinates and geometries, such as point-based locations and bounding boxes, to specify the geographic areas covered by maps. From the perspective of map image search, ISR is useful for keyword-based matching, e.g., finding map images based on the keyword of “California”, while DSR is suitable for coordinate-based matching, e.g., finding map images based on a rectangle that a user draws on a map interface. Ideally, a map image should have both ISR and DSR in its metadata, and they should be consistent with the actual geographic area covered by the map.

In reality, however, the metadata of map images are often limited. First, some map images simply do not have any ISR or DSR. Ureña-Cámara *et al.* (2019) evaluated 4,824 metadata records from a Spanish Spatial Data Infrastructure (SDI), and found that 14% of the examined records lacked spatial reference. Second, for the map images that have DSR or ISR, the provided spatial reference may not match the actual geographic area covered by the map. Renteria-Agualimpia *et al.* (2015) pointed out that the bounding box (BBox) information in DSR is often larger than the actual geographic area of a map. A global-scale survey on WMS layers also confirmed this issue and noted that many WMS layers use a global BBox by default while their map contents are about only part of the world (Gui *et al.* 2016). Figure 1 provides an example in which the content of a

WMS map image focuses on the state of Alaska while the spatial extent in its metadata covers the entire United States.

Figure 1. A map image from a WMS whose content focuses on the state of Alaska while the BBox in its metadata covers the entire United States.

One approach to addressing the issue of metadata missing or mismatching is to extract additional information from map images and enrich the existing metadata with the extracted information. Specifically, a computational model can be developed to automatically recognize the main geographic area covered by a map (e.g., Alaska) and then add relevant spatial extent and place name (in the form of a textual tag) to the metadata. Recent advancements in deep learning and computer vision offer new possibilities for accurately recognizing objects from images (LeCun *et al.* 2015). However, training a deep learning model often requires a large set of labeled data. For example, many convolutional neural networks (CNN) were trained on the ImageNet dataset which has over 14 million labeled images (Russakovsky *et al.* 2015). When it comes to recognizing geographic areas of map images, a deep learning model may need to be trained on a large number of labeled images for each geographic area to be recognized. In addition, since maps can have different colors, hues, projections, and geometric elements (e.g., points, lines, and polygons), the training dataset may need to include even more data records to reflect these diverse cartographic properties. Creating a labeled training dataset with thousands or more data records requires considerable human effort and resources.

In this work, we propose a deep learning approach with GIS-based data augmentation for enriching the metadata of map images. Compared with general data augmentation techniques that rely on existing labeled images, our approach starts from commonly available shapefiles of different geographic areas and utilizes a GIS processing package, such as GeoPandas (used in this work) or ArcPy, to generate labeled map images. Further, our approach leverages the GIS operation of map projection to project the original shapefiles into map images with different shape distortions, and includes a number of other data augmentation strategies, such as scaling, resizing, and rotating. Accordingly, our approach can automatically generate a large number of labeled map images as the training data without requiring human effort to manually label images. We use this approach to train deep learning models for enriching the metadata of map images with spatial extents and place names (in the form of textual tags). The enriched metadata can facilitate the search and discovery of map images in geoportals and other data repositories. The contributions of this paper are as follows:

- We present a deep learning approach with GIS-based data augmentation for enriching the metadata of map images. Such an approach reduces the need for manually creating a large set of labeled map images in order to train deep learning models.

- We investigate the effectiveness of individual data augmentation strategies via ablation studies, and systematically evaluate the proposed approach by using it to train multiple deep learning models and testing them on two different map image datasets.
- We share the source code of our GIS-based data augmentation, the used deep learning models, and the two test datasets of map images for supporting future research. These resources are available at: `https://doi.org/10.6084/m9.figshare.14308874`.

The remainder of this paper is organized as follows. Section 2 reviews existing studies on metadata enrichment and deep learning research related to map images. Section 3 presents the methodological details of our deep learning approach with GIS-based data augmentation. In Section 4, we evaluate the proposed approach by applying it to two different test datasets and discussing experiment results. Finally, Section 5 summarizes this work and discusses future directions.

2. Related work

The importance of metadata for discovering and reusing geospatial resources has been widely recognized by the GIScience community. Many studies have been conducted to enrich the metadata of geospatial resources hosted in geoportals, digital libraries, or other data repositories. Some studies looked into the existing metadata of a resource and proposed methods for extracting additional and often more structured information. Freire *et al.* (2011) developed a geoparsing system that can recognize place names mentioned in metadata descriptions and enrich existing metadata with recognized and geo-located place names. A similar idea of extracting place names from textual descriptions of metadata was discussed by Ureña-Cámara *et al.* (2019) in the context of examining the quality of metadata hosted in a Spanish SDI. Hu *et al.* (2015) developed a Labeled Latent Dirichlet Allocation (LLDA) based approach that can enrich the metadata of geospatial resources in a geoportal with topic categories, such as "transportation" and "health", identified from the existing textual descriptions. While these methods can extract useful information, they rely on the availability of existing metadata. In situations when little or no metadata is available, these methods cannot generate much new data.

Some other studies examined the images associated with geospatial resources. These images may be geospatial resources themselves (such as digital maps), the output of Web Map Services, thumbnails, or screenshots of geospatial datasets. One major advantage of this image-based approach is that it can be applied to situations when little or no metadata is available. In addition, this approach can help detect inconsistency between provided metadata and actual map content (Hu *et al.* 2016). Díaz *et al.* (2007) developed a software tool called *gvSIG* that can semi-automatically extract metadata from imagery and cartographic data. Their approach was based on the assumption that metadata exist in these images (e.g., in the header of a GeoTIFF image) and did not attempt to extract information from actual image content. Florczyk *et al.* (2012) proposed a method for analyzing the color features of WMS images in order to identify orthoimages from other types of WMS. Relevant textual tags, such as "orthoimage", can then be added to the metadata to facilitate the search and retrieval of these orthoimages. Evans *et al.* (2017) developed *LiveMaps*, a system that can automatically identify the geographic extents of map images from the Web by fine-tuning a ResNet model. However, their method requires manual effort to construct a labeled training dataset. Zhou *et al.* (2018) adopted a deep learning approach for classifying seven types of map images, such as "topographic map", "urban scene map", and "3D map". They manually constructed a labeled dataset with 1,812 map images, and used it to train and compare the performances of five CNN models. Yang *et al.* (2019) developed a latent-feature-based multimodality fusion model for classifying the themes of map images; however, their model was also trained on a manually labeled dataset.

Labeled training data are indispensable for many deep learning models. Depending on the complexity of the problem, the size of the training dataset can vary. For example, in the previous study by Zhou *et al.* (2018), they manually labeled 1,812 map images for classifying seven map types with the goal of including at least 200 labeled training images for each map type. While it is still possible to manually label about two thousand map images, it can be very difficult to label, e.g., 10,000 map images for the task of classifying 50 states in the U.S. (if we were to include at least 200 training maps for each state). Crowdsourcing platforms, such as Amazon's Mechanical Turk (AMT) (Paolacci *et al.* 2010), have been employed for collecting large amounts of labeled data (Russakovsky *et al.* 2015). However, running an experiment on AMT usually requires designing a Web-based user interface, validating and cleaning the collected data, and providing financial incentives (Yan *et al.* 2017, Wallgrün *et al.* 2018, Hu *et al.* 2019). These requirements increase the difficulty and cost of creating a large labeled dataset.

This work proposes a deep learning approach with GIS-based data augmentation for enriching the metadata of map images. It can be distinguished from the previous studies in the following aspects. First, we enrich the metadata of map images by examining actual map content rather than relying on existing metadata. This allows our approach to be generalized to any map images with or without metadata. For those with metadata, our approach can help detect metadata inconsistency or incompleteness. While for those without metadata, our approach can be used for adding new metadata. Second, our approach does not rely on manually labeled training data but automatically generates a large set of data from existing shapefiles. Since shapefile data are widely available nowadays, our approach can be generalized to various geographic areas and scales. Finally, much deep learning research has been conducted based on remote sensing images (Tuia *et al.* 2009, Maggiori *et al.* 2016, Zhu *et al.* 2017, Li and Hsu 2018, Marcos *et al.* 2018), and a smaller number of studies have explored the use of deep learning for analyzing map images (Duan *et al.* 2018, Zhou *et al.* 2018, Kang *et al.* 2019). Different from remote sensing images that capture real-world objects on the ground by their spectral information, map images are based on symbolic representation of geographic features and are produced using map projection, thematic rendering, and other cartographic techniques. Therefore, our work also contributes to this relatively smaller size of literature on analyzing map images using deep learning.

3. Methods

3.1. *Problem formalization*

We start by formalizing the problem of enriching the metadata of map images addressed in this work. The analysis target is a map image, which can be scanned from a paper map or directly generated in a digital format. By analyzing this map image, we aim to identify its major geographic area and enrich its metadata with two important elements: a spatial extent (in the form of a BBox) and a place name (in the form of a textual tag). Since there can be an infinite number of geographic areas at different geographic scales and extents, we assume that there exists a set of candidate geographic areas to which a map image is to be identified, e.g., one of the 50 U.S. states. Thus, we formalize the problem as below:

Given a map image m and a set of candidate geographic areas G, with $G = \{g_1, g_2, ..., g_n\}$, identify g_s from G that best matches the geographic area covered by m, and enrich the metadata of m with spatial extent e and place name t.

This problem formalization focuses on the actual content of the map image m, and does not rely on the availability of existing metadata. Once the geographic area g_s is identified, its spatial extent e and place name t can be retrieved from a geographic knowledge base, such as a gazetteer. Depending on the specific application, spatial extent e can include not only a bounding box but also the actual geometric shape of g_s; similarly, place name t does not have to be limited to one single official place name but can include multiple names, including vernacular place names, to facilitate the search of map images.

3.2. *Methodology overview*

We propose a deep learning approach with GIS-based data augmentation for enriching the metadata of map images. Unlike general images studied in computer vision (e.g., cat and dog photos), map images are about geographic areas which often have unique shapes. For example, the shape of Australia is clearly different from the shape of the United States. Given the wide availability of shapefiles for many geographic areas throughout the world, it is possible to automatically generate map images for one area using a sequence of GIS operations. These automatically generated map images can then be used as labeled data for training machine learning models.

Real-world map images are often rendered in a variety of colors and styles. The choice of color schemes or thematic styles can be application specific. However, humans with a reasonable amount of geographic knowledge can recognize the geographic area covered by a map in different colors and styles, without having to be "trained" first with these map styles and colors. This can be attributed to the ability of humans to focus on the geometric shape of the area covered by the map, rather than its particular color or style. We incorporate such a focus into the design of our approach.

Figure 2 illustrates our methodological framework, which consists of six major steps. In

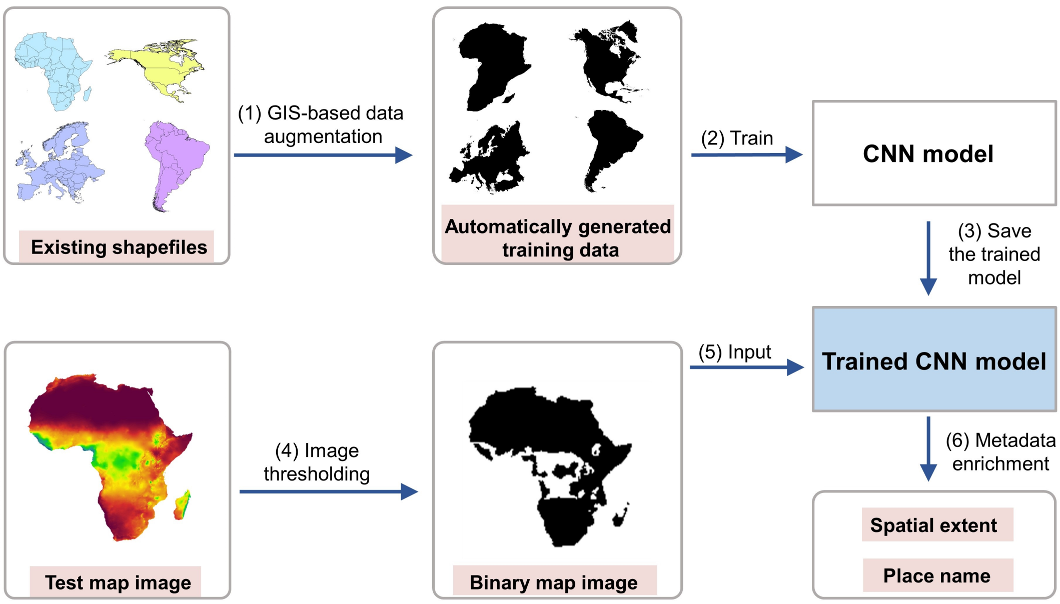

Figure 2. Overview of the methodological framework for enriching the metadata of map images.

step (1), existing shapefiles about candidate geographic areas G go through the GIS-based data augmentation process which automatically generates a training dataset containing labeled map images. In step (2), these automatically generated training data are used to train a CNN model. After the training process is finished, step (3) saves the trained CNN model for next steps. In step (4), a test map image m, whose geographic area is to be identified, goes through an image thresholding process that converts the original map image into a black-and-white binary image. Such a conversion process removes the color and style information of the map image and lets the model focus on the shape of the covered geographic area. In step (5), the converted binary image is used as the input of the trained CNN model. Finally, in step (6), the trained CNN model recognizes the geographic area g_s from the map image, and the spatial extent e and place name t are retrieved from the used geographic knowledge base. In the following, we present details of this methodological framework.

3.3. *GIS-based data augmentation*

The step of GIS-based data augmentation leverages existing shapefiles of different geographic areas and utilizes a GIS processing package to automatically generate a training dataset. Unlike typical data augmentation approaches that modify existing labeled images (Géron 2019), our approach generates map images directly from shapefiles. Specifically, our GIS-based data augmentation involves the following major strategies:

(1) **Map projection.** Map projection is an essential component of any map. It transforms geographic features from the surface of a 3D globe onto a 2D plane. Different map projections can cause different shape distortions of the same geographic area. Figure 3 shows an example of the contiguous U.S. under the Plate Carrée and Albers Equal Area projections respectively. Since there exist a large number of map projections, one can select those that are often applied to a geographic area. Here, we recommend using the Projection Wizard tool developed by Šavrič *et al.* (2016), which can suggest suitable map projections based on a target geographic area. In addition, one can include some commonly used map projections, such as Web Mercator, although they may not be considered as suitable for a target area (e.g., the entire world) from a cartographic perspective.

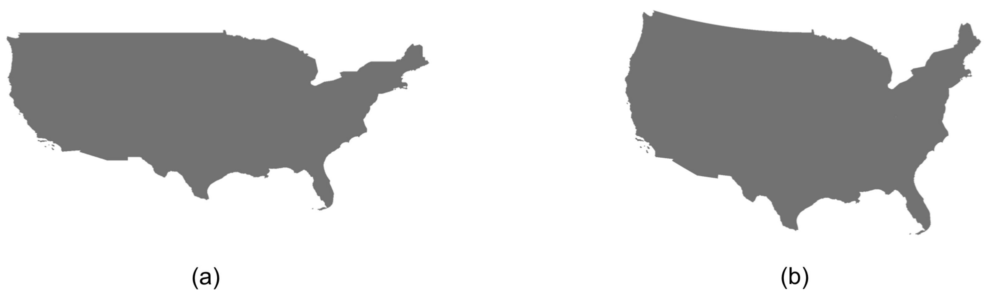

Figure 3. Different shapes of the contiguous U.S. under two map projections: (a) Plate Carrée (WGS84); (b) Albers Equal Area.

(2) **Scaling.** Real-world maps may depict the same geographic area at different scales. To simulate a similar visual effect, we scale one map image both inward

and outward, which makes the covered geographic area look smaller or larger in the map image. The images are scaled both inward and outward by 5%, 10%, 15%, and 20%. We limit the scaling extent within 20% to prevent excessive transformations that can lead to a large loss of shape information.

(3) **Resizing.** Map images from the Web and other sources may be squeezed vertically or horizontally. For example, a map image requested from a WMS can have arbitrarily-defined width and height depending on the parameters used in the request. Figure 4(a) shows a squeezed California map image from a WMS, and Figure 4(b) shows a California map with a more proper width-to-height ratio. To accommodate these varied ratios, we resize the map images with four different width-to-height ratios which are 50%, 75%, 125%, and 150%.

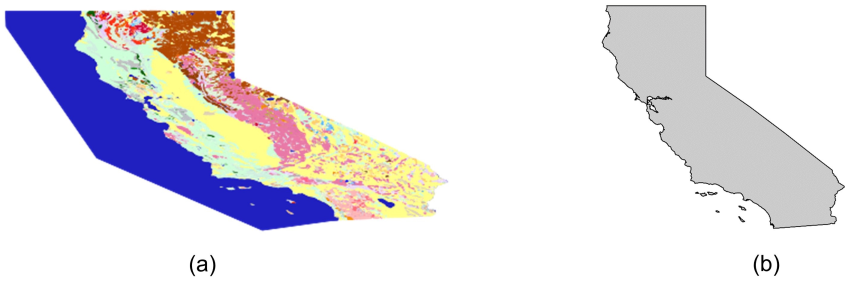

Figure 4. An example of a squeezed map image: (a) California map requested from a WMS; (b) California map with a more proper width-to-height ratio.

(4) **Rotation.** The geographic area shown on a map image may also be rotated. For example, a paper map may be rotated slightly during the scanning process which results in a rotated map image. Here, we rotate map images by angles from 0 to 30 degrees with an increment of 10 degrees. The rotation is performed around the center of the image in both clockwise and counterclockwise directions.

(5) **Noise addition.** Map images from different sources on the Web or in libraries may not be of the highest quality. Noise can be introduced when paper maps are scanned or when a map image is further processed. Here, we add Gaussian noise with a mean of 0 and a standard deviation of 0.1 (Krizhevsky *et al.* 2012, Shijie *et al.* 2017) to the map images to simulate noise effect.

(6) **Blurring.** Similar to noise addition, this strategy aims to reduce the acuity of map images to make them similar to those real-world map images with lower resolutions. Particularly, we apply a convolution function with the widely used Gaussian kernel (Hussain *et al.* 2017, Casado-García *et al.* 2019) to process map images. The Gaussian kernel is defined by a random variance σ chosen from 0 to 3.0, and the kernel radius r is generated by $r = 4 \times \sigma$ (Hussain *et al.* 2017).

Figure 5 illustrates the results after applying each of the six strategies to a map of North America. Note that some of these data augmentation strategies (e.g., scaling and resizing) can generate map images with different width and height, whereas many existing CNN models require the input image to have a square shape with a fixed size. Therefore, we also apply image processing operations such as padding and cropping to ensure that the generated map images have the same width and height. The actual sizes of the generated training images vary depending on the target geographic area and the applied augmentation strategies, but their width and height are always the same. When these

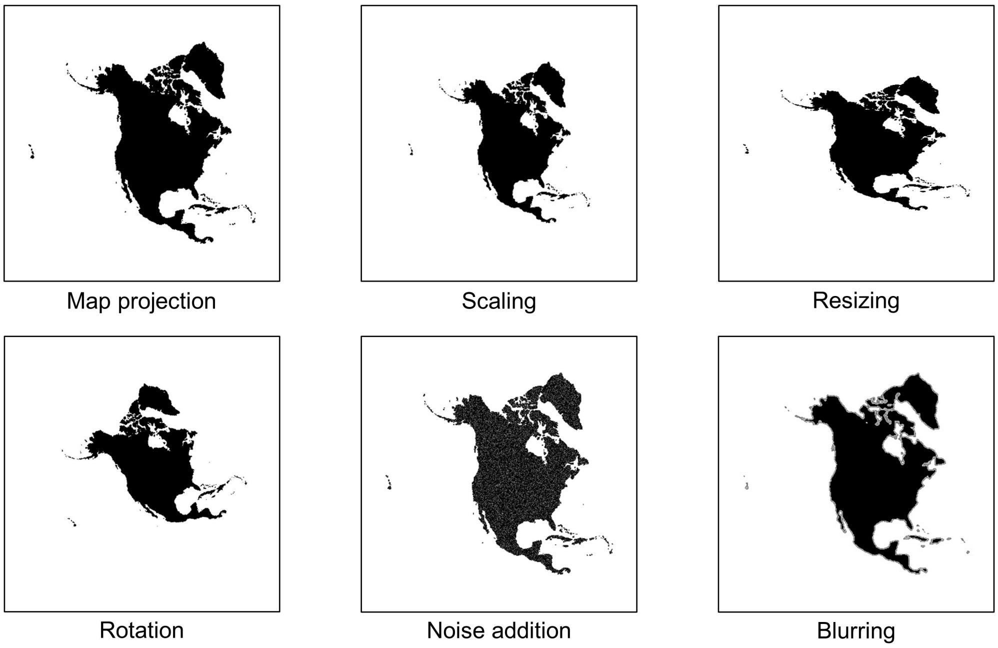

Figure 5. Results of individual data augmentation strategies applied to the same map.

generated map images are used to train a CNN model, they are dynamically resized to the input size required by the model (e.g., 224 by 224).

Figure 6 shows our GIS-based data augmentation step by step. Let n represent the

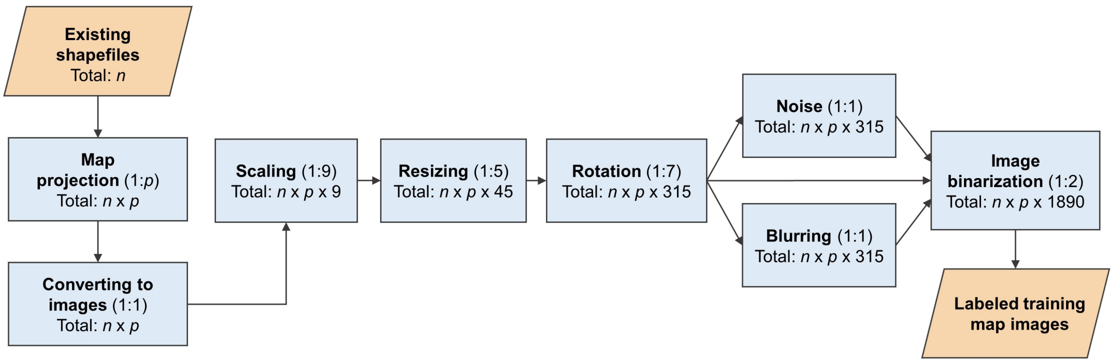

Figure 6. GIS-based data augmentation for generating labeled training map images.

number of input shapefiles (one shapefile per geographic area), and p represent the number of selected projections for each shapefile. The n input shapefiles are first *projected* using the p selected map projections. The projected shapefiles (vector data) are then converted to map images (raster data). In the following steps, one input image is first *scaled* to 9 new images, then *resized* into 45 images, and *rotated* to 315 images. Next, each of the 315 images will be used for generating two additional new images: one for *blurring* and one with *noise addition*. For the final step, we convert each of the generated images into two binary images: one with black foreground and white background, and the other with white foreground and black background. We generate these black-and-white binary images to help the trained model focus on the shapes of target geographic areas. We generate two binary images instead of one, because the image thresholding step, which will be applied to a test map (see Section 3.5), can produce both types of binary images depending on the color scheme of the map. In total, our GIS-based data augmentation will generate ($p \times 1,890$) labeled training images per shapefile. If we choose 5 map projections (i.e., $p = 5$), this approach can automatically generate 9,450 labeled map images for one geographic area. A GIS processing package, such as GeoPandas or ArcPy, along with an image processing package, such as OpenCV, can be utilized to implement this GIS-based data augmentation.

3.4. *CNN models*

With automatically generated map images, a CNN model can be trained to recognize the geographic area from a real-world map. Many CNN models have been proposed in the literature, and three are selected and used in this work, which are AlexNet, Inception-v3, and ResNet. These three models were all champions in previous ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competitions, and using three different models allows us to test the generalizability of the proposed GIS-based data augmentation, rather than relying on the performance of one particular model.

- **AlexNet**. As one of the early deep CNNs, AlexNet was the winning model of ILSVRC 2012 (Krizhevsky *et al.* 2012). It is still widely used in many applications nowadays as a baseline for image classification and object detection. The model consists of 5 convolutional layers, 3 maxpooling layers, and 3 fully connected layers.
- **Inception-v3**. Inception-v3 is an improved version of GoogLeNet (Szegedy *et al.* 2015) which won ILSVRC 2014. GoogLeNet introduced an innovative *inception module* in which some layers run in parallel. Inception-v3 improves the original GoogLeNet by replacing the 5x5 convolutional layer with three 3x1 layers and adding factorized convolutions. This improvement leads to much lower error rate. The Inception-v3 model used in this work has a depth of 42 layers.
- **ResNet**. The ResNet model developed by He *et al.* (2016) was the winning model of ILSVRC 2015. ResNet reduces the errors associated with very deep neural networks by introducing a building block called *residual block*. ResNet can capture more high-level compositional features and detect fine-grained objects from images. In this work, we use the version of ResNet with 50 layers (ResNet-50), which requires the smallest floating point operations.

For all three models, we replace their final output layer with a fully connected layer consisting of n neurons, where n is the number of candidate geographic areas. While our experiments focus on these three CNNs, other models can also be trained on the generated map images.

3.5. *Image thresholding*

A map image can be rendered in many possible colors and cartographic styles, depending on the specific application and the design choices made by the cartographer. To help a model focus on the shape of a geographic area rather than map colors or textures, we use the technique of *thresholding* to convert an original map image whose metadata need to be enriched into a black-and-white binary image. This is shown as step (4) in the methodological framework in Figure 2.

Thresholding is a simple but effective way to segment the boundary of a target object from its background (Sezgin and Sankur 2004). Existing thresholding algorithms can be classified into two categories: (1) histogram-based thresholding (Zack *et al.* 1977, Jiao *et al.* 2006), in which a histogram of pixel intensity is created and used for deciding the threshold; (2) local thresholding (Yen *et al.* 1995, Ridler *et al.* 1978, Li and Lee 1993), in which the characteristics of neighboring pixels are analyzed to process each pixel. To identify a suitable thresholding method, we tested seven different algorithms on a set of WMS map images and manually examined their results. An example is shown in Figure 7. We eventually chose the *Triangle* thresholding method since it best separates the covered

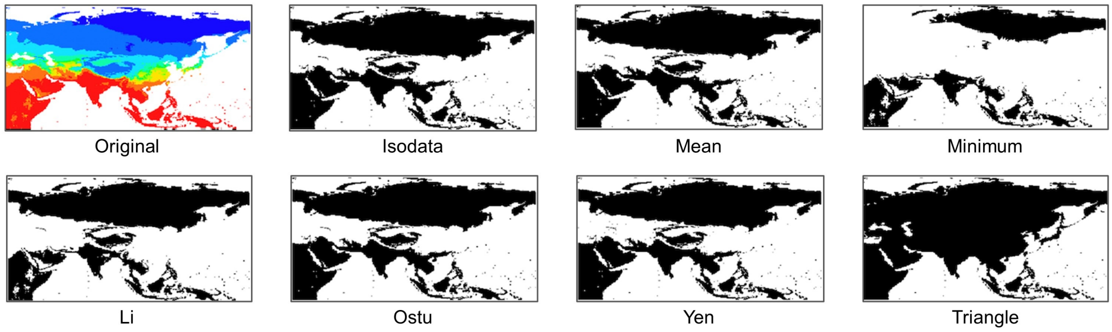

Figure 7. Results of seven different thresholding methods applied to the same map image.

geographic area from its background. We also tested the possibility of converting maps to grey-scale images rather than black-and-white ones. We found that grey-scale images cannot clearly show the boundaries of some geographic areas when they have colors similar to their background, although grey-scale images can also reduce the distraction of colors on the model.

In many cases, the *Triangle* thresholding method will render the main geographic area covered by a map as black and the background as white. This happens when the main geographic area is rendered with rich colors while the background has bland colors. In some other cases, however, it can also produce maps with the main geographic area as white and the background as black. This happens when the background of a map has rich colors. For example, an ocean temperature map of Australia may use a bland color for the continent and rich colors to represent the temperature variation in the ocean. Our GIS-based data augmentation handles this by generating both types of binary images for one geographic area.

4. Experiments and results

In this section, we evaluate the performance of our approach in two sets of experiments using map images from different data sources and at different geographic scales. In both experiment sets, the CNN models are trained or fine-tuned using the map images generated by our GIS-based data augmentation and are then tested on real-world map images. Thus, no map image in the test dataset is used in the training process. This also means we cannot use some validation methods, such as k-fold cross-validation, which assume that the training and validation data are from the same or similar datasets. Here, the training data and test data are different, since the former are generated map images while the latter are real-world maps.

4.1. *Experiment set I: WMS map images at the continental scale*

In the first set of experiments, we evaluate our approach on a test dataset of map images collected from a global WMS quality survey (Gui *et al.* 2016). This test dataset contains 800 map images in total, with 100 maps for each of the seven continents and 100 maps for the entire world. Thus, the classification task based on this dataset has 8 categories. A sample of these map images is shown in Figure 8.

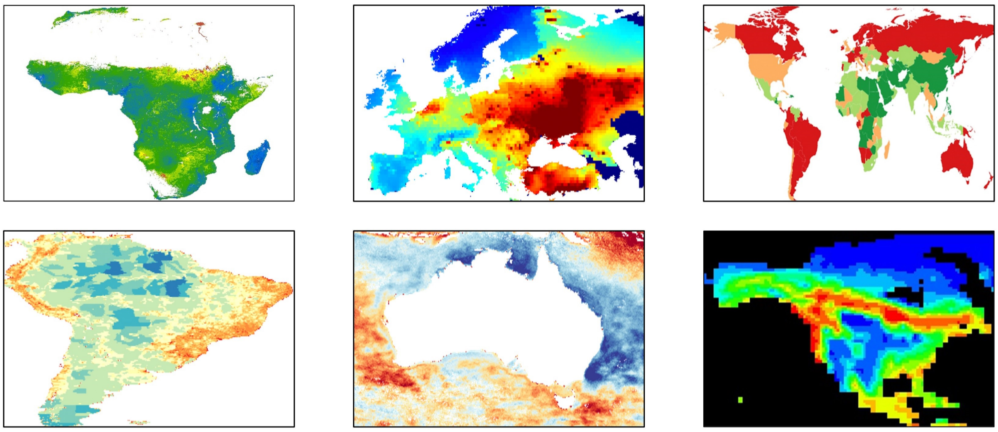

Figure 8. A sample of map images in the continent dataset.

We use our GIS-based data augmentation to generate labeled map images as the training data. Shapefiles of the seven continents and the entire world are obtained from Esri Data & Maps (shapefiles from other sources, such as Natural Earth, can be used as well). For each continent, we select five map projections using the Projection Wizard tool (Šavrič *et al.* 2016). Whenever possible, we try to include one projection from each of the three types: equal-area, conformal, and equidistant, and we also include Web Mercator and Plate Carrée (WGS84) projections. For the entire world, we select six commonly used world map projections. Table 1 summarizes these map projections. In total, 77,490 labeled map images are generated through the GIS-based data augmentation, with 9,450 map images ($1,890 \times 5$ projections) for each continent and 11,340 images ($1,890 \times 6$ projections) for the world.

Next, we use the generated map images to train CNN models. We leverage the pre-trained AlexNet, Inception-v3, and ResNet models based on the ImageNet dataset, and fine-tune them using the map images generated from our GIS-based data augmentation. Pre-trained models have already learned patterns typically from a large benchmark dataset (e.g., 14 million images in the ImageNet dataset), and fine-tuning them avoids learning everything from scratch. Such a transfer learning process reduces the training time, and can often achieve very good performances on similar tasks (Géron 2019). Meanwhile, considering that our task of classifying black-and-white map images is also different from a task on natural RGB images, we train an AlexNet model completely from scratch using the generated map images. Table 2 shows the classification accuracy of the tested models, which is calculated as the number of correctly classified maps divided by the total number of test maps. We also train and test the same models without using the GIS-based data augmentation to evaluate the effectiveness of our approach.

As shown in Table 2, the models trained on the map images generated by our GIS-based data augmentation all substantially outperform the models without data augmentation. This result demonstrates that our approach is overall effective and that it can

Table 1.: Map projections used for the seven continents and the world (note that Web Mercator and Plate Carrée are added to each category).

	Map Projections
Africa	*Lambert azimuthal equal-area, Equatorial Stereographic, Equidistant cylindrical*
Antarctica	*Lambert azimuthal equal-area, Azimuthal equidistant, Polar stereographic*
Asia	*Albers equal-area conic, Lambert conformal conic, Equidistant conic*
Europe	*Albers equal-area conic, Lambert conformal conic, Equidistant conic*
North America	*Albers equal-area conic, Lambert conformal conic, Equidistant conic*
Oceania	*Albers equal-area conic, Lambert conformal conic, Equidistant conic*
South America	*Oblique Lambert azimuthal equal-area, Oblique Stereographic, Equidistant conic*
World	*Robinson, Natural Earth, Mollweide, Wagner IV*

Table 2.: Classification accuracy of the tested models on the WMS continent dataset.

	AlexNet (trained)	AlexNet (fine-tuned)	Inception-v3 (fine-tuned)	ResNet (fine-tuned)
GIS-based data augmentation	**0.247**	**0.501**	**0.580**	**0.532**
No data augmentation	0.040	0.014	0.013	0.011

be generalized to different CNN model architectures. The models trained without data augmentation have very low classification accuracy. This can be attributed to the insufficient number of training data, i.e., there is only one training map image for each category generated from the default shapefile. This result further highlights the challenge of using deep learning models when small or no labeled data is available, and our work is an effort toward addressing this challenge. By comparing the AlexNet model trained from scratch with the other three models, we can see that the pre-trained and fine-tuned models have better performances. Although those three models were pre-trained on RGB images, it is possible that they have learned to detect some patterns, such as lines and shapes, that can still be applied to this task. The highest accuracy 0.580 is obtained by Inception-v3. While this is not a very high score, it is achieved using only the map images automatically generated from shapefiles.

While our GIS-based data augmentation is shown to be effective overall, which strategies work better or are they equally effective? To answer this question, we conduct an ablation study by removing one data augmentation strategy at a time and keeping the rest strategies unchanged. We focus on the three fine-tuned models shown to be more effective in the previous experiment. Table 3 presents the result of this ablation study.

Two observations are obtained. First, when either *projection*, *scaling*, or *resizing* is removed from the data augmentation process, the accuracy of the three models all decreases. This result shows the effectiveness of these three data augmentation strategies. Particularly, we see a decrease of about 0.2 in accuracy (or a relative decrease of about

Table 3.: Classification accuracy of the tested models on the WMS continent dataset when different augmentation strategies are removed.

	AlexNet (fine-tuned)	Inception-v3 (fine-tuned)	ResNet (fine-tuned)
All strategies	0.501	0.580	0.532
No projection	0.321	0.371	0.352
No scaling	0.462	0.523	0.368
No resizing	0.427	0.572	0.457
No rotation	**0.628**	**0.689**	**0.775**
No noise	0.586	0.533	0.610
No blurring	0.584	0.547	0.600

40%) when *projection* is removed. This result demonstrates the critical role of map projection in helping a CNN recognize the geographic area covered by a map. Second, removing one of the other three strategies, namely *rotation*, *noise*, and *blurring*, increases the performance of the models in many cases. This result is surprising and suggests that the other three strategies might not be effective. With curiosity, we test the performance of the models by removing these three strategies all together. The classification accuracy scores, however, are 0.513, 0.547, and 0.596 for AlexNet, Inception-v3, and ResNet respectively, suggesting that removing all three strategies does not achieve the highest accuracy either. This might be attributed to the largely reduced number of training data records when all three strategies are removed. Based on Table 3, the highest accuracy is 0.775, which is achieved by ResNet when the strategy of *rotation* is removed.

Lastly, we compare our approach with *LiveMaps*, a system similar to our work that can identify the geographic viewpoint (similar to spatial extent) of a given map image (Evans *et al.* 2017). LiveMaps also uses a pre-trained ResNet model. The major difference between their approach and ours is that they fine-tune the ResNet model using real map images crawled from the Web rather than map images automatically generated from shapefiles. However, their approach of crawling training map images from the Web is not completely automatic, as manual effort is required to examine the crawled images and remove noise. In addition, their approach assumes the availability of relevant map images on the Web for a particular geographic area, and cannot be applied to those areas that do not have many map images online. To implement *LiveMaps*, we use Google Images Download (a Python-based Web crawler) to crawl map images about the world and the seven continents from Google Image Search, and manually clean the crawled map images which are then used to train the *LiveMaps* model. The trained model is applied to the same test dataset of WMS continent map images. It achieves a classification accuracy of 0.740, which is higher than our initial score 0.580 when all data augmentation strategies are used but lower than our best score 0.775 achieved by ResNet.

4.2. *Experiment set II: Web map images at the state scale*

In the second set of experiments, we evaluate our approach on a test dataset of map images about the 50 U.S. states. These map images were collected from the Web using the Google Images Download crawler with keywords in the format of "<StateName>

map" (e.g., "California map"). The first 200 search results were downloaded for each state, and we manually examined the result to remove noise images with the goal of keeping 50 images for each of the 50 states. In total, this dataset has 2,500 map images. The classification task based on this dataset has 50 categories. A sample of these state-level map images is provided in Figure 9.

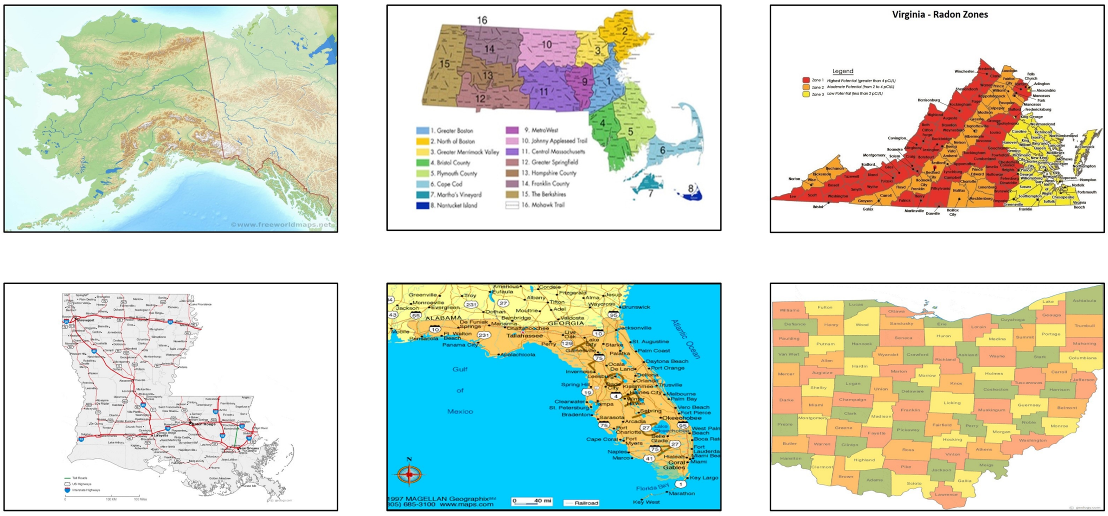

Figure 9. A sample of map images in the U.S. state dataset.

We generate training map images using the shapefiles of the U.S. states from Esri Data & Maps. For each state, we adopt five map projections, which are Albers equal-area conic, Lambert conformal conic, oblique azimuthal equidistant, UTM (based on the zone of each state), and Plate Carrée (since some maps directly visualize states based on their WGS84 coordinates). In total, our GIS-based data augmentation generates 472,500 map images, with 9,450 images (1,890 images $\times$ 5 projections) for each of the 50 states.

The generated map images are then used to fine-tune the three pre-trained CNN models. Similar to the previous set of experiments, we also train an AlexNet model from scratch using the generated map images. Table 4 shows the classification accuracy of the tested models.

Table 4.: Classification accuracy of the tested models on the US state dataset.

	AlexNet (trained)	AlexNet (fine-tuned)	Inception-v3 (fine-tuned)	ResNet (fine-tuned)
GIS-based data augmentation	**0.299**	**0.436**	**0.431**	**0.330**
No data augmentation	0.023	0.008	0.013	0.030

Similar to the results of previous experiments, using the GIS-based data augmentation substantially improves the accuracy of all tested models. The highest accuracy 0.436 is achieved by AlexNet. While it is only a fair score, classifying these state map images is highly difficult. This difficulty can be seen in two aspects. First, there are 50 classes in this task and a random classification on this dataset should achieve an average accuracy of only 0.020. Second, the map images in this dataset have highly complex visual elements, colors, cartographic styles, and even distortions. For example, some map images in Figure 9 contain multiple types of map symbols, customized legends, and various text annotations, which make it difficult to generate map images with similar visual features. It is also difficult for the image thresholding method to derive accurate boundaries of the target geographic areas.

We further conduct an ablation study based on the three pre-trained models to examine the effectiveness of different data augmentation strategies. The results are shown in Table 5. Different from the result of experiment set I, the best performance is achieved

Table 5.: Classification accuracy of the tested models on the U.S. state dataset when different augmentation strategies are removed.

	AlexNet (fine-tuned)	Inception-v3 (fine-tuned)	ResNet (fine-tuned)
All strategies	**0.436**	**0.431**	**0.330**
No projection	0.206	0.203	0.214
No scaling	0.322	0.206	0.181
No resizing	0.250	0.219	0.205
No rotation	0.365	0.297	0.274
No noise	0.329	0.310	0.283
No blurring	0.335	0.322	0.232

by using all data augmentation strategies, and removing any strategy reduces the performance of the models. This performance difference suggests that the effectiveness of a data augmentation strategy may depend on the characteristics of the target map images. Most maps from the WMS continent dataset have little rotation, noise, or blurring effect, since they are images retrieved from WMS services. As a result, the corresponding data augmentation strategies become less effective and may even reduce the performance of the trained model. By contrast, the US state maps crawled from the Web have diverse characteristics in terms of their rotations, noise, and image quality. Thus, the data augmentation strategies of *rotation*, *noise*, and *blurring* may become more effective as shown in the experiment results. Finally, in this set of experiments, we are unable to compare our approach with *LiveMaps*, because it is highly labor-intensive to construct a training dataset by crawling and manually examining map images for 50 states. In addition, most state map images available from the Web (at least in the first 200 search results) are already used in the test dataset and therefore should not be used as training data.

4.3. *Discussion*

The two sets of experiments based on different datasets and different CNN models enhance our understanding of the proposed deep learning approach with GIS-based data augmentation. Overall, the experiment results suggest that our approach can substantially boost the performance of CNN models in their capability of recognizing the covered geographic areas of map images. Meanwhile, the effectiveness of individual data augmentation strategies seems to be associated with the characteristics of map images to which the trained models will be applied. Three strategies, namely *projection*, *scaling*, and *resizing*, have consistently increased the performance of the trained models on the two datasets, while the strategies of *rotation*, *noise*, and *blurring* are effective when the test map images share similar characteristics simulated by these strategies.

Without requiring manual effort for data labeling, the proposed approach can be used for enriching the metadata of map images, particularly spatial extents and place names, when there are limited human or financial resources for creating a labeled training dataset manually. In addition, even when we have the necessary resources to do so, the ability to automatically generate map images from shapefiles reduces the time needed to create a large labeled dataset. With the wide availability of shapefiles, our approach can be generalized to many geographic areas at different scales throughout the world.

The experiments also show that our approach so far has only fair classification accuracy. It is worth noting that all training data are generated from shapefiles through the GIS-based data augmentation, and the trained models have never seen any real-world map image. This is different from typical machine learning tasks in which the training and test data are largely similar (or the training and test data can be considered as sampled from the same distribution). The two classification tasks in our experiments are thus very difficult for the trained models, due to the inevitable difference between the generated training map images and the test real-world maps. This is illustrated in Figure 10, which shows (a) a real-world map of New York State, (b) its black-and-white version after image thresholding, and (c) a map image generated from shapefile. While our generated

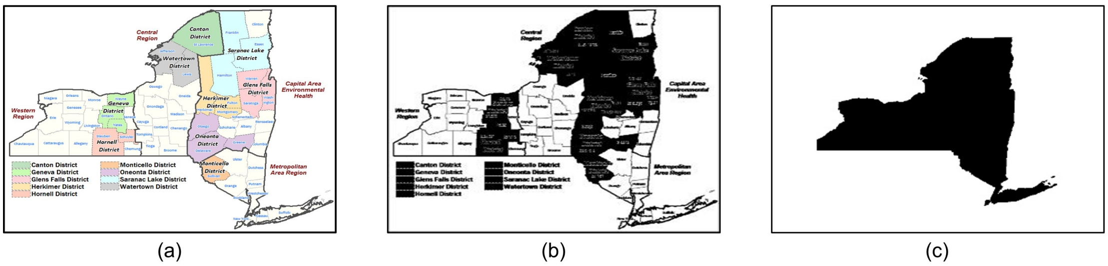

Figure 10. Difference between a real-world map of New York State and a generated map: (a) the real-world map; (b) the black-and-white map after image thresholding; (c) a map generated from shapefile.

map image well captures the shape of New York State, it nevertheless does not have the legend, text annotations, and county boundaries in the real-world map. To further increase the similarity between generated map images and real-world maps, we could consider automatically creating and adding those map elements to a generated map image, or alternatively, removing them from the real-world map.

5. Conclusions and future work

In this paper, we presented a deep learning approach with GIS-based data augmentation for enriching the metadata of map images. Compared with general data augmentation techniques, our approach can automatically generate labeled training map images directly from shapefiles leveraging a GIS processing package, such as GeoPandas. Our approach integrates six data augmentation strategies, including map projection, scaling, resizing, rotation, noise addition, and blurring. We used our GIS-based data augmentation to automatically generate labeled map images for training and fine-tuning multiple CNN models, and tested them on two datasets: a set of WMS images containing world and continent maps and a set of U.S. state map images crawled from the Web. Experiment results suggest that the proposed approach can substantially increase the performance of CNN models on recognizing geographic areas from maps, and can help enrich the metadata of map images, particularly spatial extents and place names, without requiring manual data labeling. This work is a step toward addressing the challenge of obtaining a large set of labeled data often needed for training deep learning models.

Our proposed approach also has potential for helping a geographic information system become more intelligent. For example, one task that a GIS user encounters from time to time is to assign a spatial reference system to a geospatial dataset whose spatial reference is missing (e.g., a shapefile that lacks the *.prj file). To complete such a task, one needs to select a suitable coordinate system for the target geographic area often from hundreds of possible coordinate systems. Our approach offers the possibility of enabling a GIS to automatically recognize the geographic area of the data based on its shape and recommend a small number of suitable coordinate systems to the GIS user.

This work could be extended in several directions. First, we can improve the proposed methodological framework by enhancing its existing steps or adding new steps. For example, image thresholding is an important step that directly affects the extracted shape of the covered geographic area. While we have experimented with seven different thresholding methods, other techniques could be explored as well. We could also add new steps to remove certain map elements from visually complex maps to help the trained models focus on the shape of the geographic area. Second, a map image collection may contain noise images that do not belong to any of the classification categories, and ideally, a trained model would identify those noise images. This is a non-trivial research problem called *open set recognition* (Scheirer *et al.* 2012), and we made a preliminary exploration on this problem in Appendix A. We found a trade-off between noise image identification and classification accuracy, but further investigation is needed on this topic. Finally, while this work has largely focused on shapes in order to identify the main geographic area of a map, future work could explore the possibility of generating map images with colors and textures similar to those of real-world maps. For example, graduated color maps could be generated based on selected attributes of a shapefile and pre-defined cartographic principles. Those generated map images can then be used to train deep learning models for recognizing thematic map types based on, e.g., varied colors in different regions. By exploring these and other directions, we can build more intelligent AI models to facilitate the extraction and use of information from large numbers of maps.

Data and code availability statement

The data and code that support this work are available on 'figshare.com' with the identifier: `https://doi.org/10.6084/m9.figshare.14308874`

Acknowledgement

Dr. Zhipeng Gui acknowledges support from the National Natural Science Foundation of China (No. 41971349).

References

Bendale, A. and Boult, T.E., 2016. Towards open set deep networks. *In*: *Proceedings of the IEEE conference on computer vision and pattern recognition* IEEE, 1563–1572.

Bernard, L., *et al.*, 2005. The European geoportal—-one step towards the establishment of a European Spatial Data Infrastructure. *Computers, environment and urban systems*, 29 (1), 15–31.

Casado-García, Á., *et al.*, 2019. CLoDSA: a tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. *BMC bioinformatics*, 20 (1), 323.

Díaz, L., *et al.*, 2007. Semi-automatic metadata extraction from imagery and cartographic data. *In*: *2007 IEEE International Geoscience and Remote Sensing Symposium* IEEE, 3051–3052.

Duan, W., *et al.*, 2018. Automatic generation of precisely delineated geographic features from georeferenced historical maps using deep learning. *In*: *Proceedings of the AutoCarto* ACM.

Evans, M.R., *et al.*, 2017. LiveMaps: Learning Geo-Intent from Images of Maps on a Large Scale. *In*: *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* ACM, p. 31.

Florczyk, A.J., *et al.*, 2012. Identifying orthoimages in web map services. *Computers & geosciences*, 47, 130–142.

Freire, N., *et al.*, 2011. A metadata geoparsing system for place name recognition and resolution in metadata records. *In*: *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries* ACM, 339–348.

Géron, A., 2019. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Gui, Z., *et al.*, 2016. Global-scale resource survey and performance monitoring of public OGC web map services. *ISPRS International Journal of Geo-Information*, 5 (6), 88.

Gui, Z., *et al.*, 2013. A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. *International Journal of Geographical Information Science*, 27 (6), 1109–1132.

He, K., *et al.*, 2016. Deep residual learning for image recognition. *In*: *Proceedings of the IEEE conference on computer vision and pattern recognition* IEEE, 770–778.

Hu, K., *et al.*, 2016. Content-based discovery for web map service using support vector machine and user relevance feedback. *PloS one*, 11 (11), e0166098.

Hu, Y., Deng, C., and Zhou, Z., 2019. A Semantic and Sentiment Analysis on Online Neighborhood Reviews for Understanding the Perceptions of People toward Their Living Environments. *Annals of the American Association of Geographers*, 109 (4), 1052–1073.

Hu, Y., *et al.*, 2015. Metadata topic harmonization and semantic search for linked-data-driven geoportals: A case study using ArcGIS Online. *Transactions in GIS*, 19 (3), 398–416.

Hussain, Z., *et al.*, 2017. Differential data augmentation techniques for medical imaging classification tasks. *In*: *AMIA Annual Symposium Proceedings*, Vol. 2017 American Medical Informatics Association, p. 979.

Jiao, S., Li, X., and Lu, X., 2006. An improved Ostu method for image segmentation. *In*: *2006 8th international Conference on Signal Processing*, Vol. 2 IEEE.

Kang, Y., Gao, S., and Roth, R.E., 2019. Transferring multiscale map styles using generative adversarial networks. *International Journal of Cartography*, 1–27.

Krizhevsky, A., Sutskever, I., and Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *In*: *Advances in neural information processing systems* Neural Information Processing Systems Foundation Inc., 1097–1105.

LeCun, Y., Bengio, Y., and Hinton, G., 2015. Deep learning. *nature*, 521 (7553), 436.

Li, C.H. and Lee, C., 1993. Minimum cross entropy thresholding. *Pattern recognition*, 26 (4), 617–625.

Li, W. and Hsu, C.Y., 2018. Automated terrain feature identification from remote sensing imagery: a deep learning approach. *International Journal of Geographical Information Science*, 1–24.

Lutz, M. and Klien, E., 2006. Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science*, 20 (3), 233–260.

Maggiori, E., *et al.*, 2016. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55 (2), 645–657.

Marcos, D., *et al.*, 2018. Land cover mapping at very high resolution with rotation equivariant CNNs: Towards small yet accurate models. *ISPRS journal of photogrammetry and remote sensing*, 145, 96–107.

Paolacci, G., Chandler, J., and Ipeirotis, P.G., 2010. Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5 (5), 411–419.

Renteria-Agualimpia, W., *et al.*, 2015. Identifying geospatial inconsistency of web services metadata using spatial ranking. *Earth Science Informatics*, 8 (2), 427–437.

Ridler, T., Calvard, S., *et al.*, 1978. Picture thresholding using an iterative selection method. *IEEE trans syst Man Cybern*, 8 (8), 630–632.

Russakovsky, O., *et al.*, 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115 (3), 211–252.

Šavrič, B., Jenny, B., and Jenny, H., 2016. Projection Wizard–An online map projection selection tool. *The Cartographic Journal*, 53 (2), 177–185.

Scheirer, W.J., *et al.*, 2012. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35 (7), 1757–1772.

Sezgin, M. and Sankur, B., 2004. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13 (1), 146–166.

Shijie, J., *et al.*, 2017. Research on data augmentation for image classification based on convolution neural networks. *In*: *2017 Chinese automation congress (CAC)* IEEE, 4165–4170.

Szegedy, C., *et al.*, 2015. Going deeper with convolutions. *In*: *Proceedings of the IEEE conference on computer vision and pattern recognition* IEEE, 1–9.

Tuia, D., *et al.*, 2009. Active learning methods for remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 47 (7), 2218–2232.

Uhl, J., *et al.*, 2018. Map archive mining: visual-analytical approaches to explore large historical map collections. *ISPRS International Journal of Geo-Information*, 7 (4), 148.

Ureña-Cámara, M.A., *et al.*, 2019. A method for checking the quality of geographic metadata based on ISO 19157. *International Journal of Geographical Information Science*, 33 (1), 1–27.

Wallgrün, J.O., *et al.*, 2018. GeoCorpora: building a corpus to test and train microblog geoparsers. *International Journal of Geographical Information Science*, 32 (1), 1–29.

Wallis, J.C., *et al.*, 2010. Digital libraries for scientific data discovery and reuse: from vision to practical reality. *In*: *Proceedings of the 10th annual joint conference on Digital libraries* ACM, 333–340.

Yan, B., *et al.*, 2017. From itdl to place2vec: Reasoning about place type similarity and relatedness by learning embeddings from augmented spatial contexts. *In*: *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* ACM, p. 35.

Yang, Z., *et al.*, 2019. A Latent Feature-Based Multimodality Fusion Method for Theme Classification on Web Map Service. *IEEE Access*, 8, 25299–25309.

Yen, J.C., Chang, F.J., and Chang, S., 1995. A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing*, 4 (3), 370–378.

Zack, G., Rogers, W., and Latt, S., 1977. Automatic measurement of sister chromatid exchange frequency.. *Journal of Histochemistry & Cytochemistry*, 25 (7), 741–753.

Zhou, X., *et al.*, 2018. Deep Convolutional Neural Networks for Map-Type Classification. *arXiv preprint arXiv:1805.10402*.

Zhu, X.X., *et al.*, 2017. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5 (4), 8–36.

Appendix A. Exploring the practicality of identifying noise images

A dataset may contain noise images that do not belong to any of the pre-defined categories. Existing CNN models typically use a softmax layer for the final output, and a noise image will always be classified into one of the pre-defined categories. Here, we explore the practicality of automatically identifying noise images. This is a non-trivial research problem called *open set recognition* (Scheirer *et al.* 2012), and we only make a preliminary exploration here.

The experiments are based on the first dataset of WMS continent map images, which is an easier dataset than the second one. With the existing 800 maps in this dataset, we add 100 WMS map images as noise maps, which do not belong to any of the eight categories. Examples of these maps are shown in Figure A1.

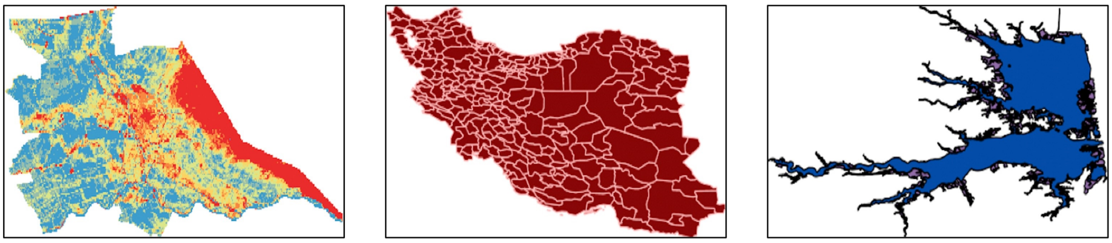

Figure A1. Some examples of the noise map images added to the test dataset.

For the CNN model, we focus on ResNet which has achieved the best performance on this dataset based on our previous experiments. To help the model detect noise images, one straightforward approach is to add training images that do not belong to any of the eight categories, and then train the model using these noise images along with the existing training data. In other words, this approach does not change the major structure of the model, but adds a noise category in the training data. However, it is difficult, if not impossible, to obtain training images that can represent all noise images, since any image that is not a continent map is a noise image. Thus, the trained model will become biased toward the particular set of noise images used for training.

Here, we use a method proposed by Bendale and Boult (2016), in which the final softmax layer is replaced by an openmax layer that transforms the original output from n categories to $n+1$ categories using a fitted Weibull distribution. In other words, this approach does not change the existing training data but alters the structure of the CNN model, thereby sidestepping the bias introduced by a particular set of noise images. We compare the result of this openmax approach with that of the original ResNet model. The comparison result is shown in Table A1.

We examine the result based on two questions. First, to what degree can we successfully detect noise map images using the openmax layer? Second, how does the classification accuracy on the existing eight categories change? For the first question, the original ResNet model with softmax layer fails to detect any noise map images in the test dataset, as shown in the last row of Table A1. As discussed previously, this result is expected due to the use of a softmax layer as the output. By contrast, the ResNet model with an openmax layer correctly detects 65% of the noise images. For the second question, the classification accuracy of the ResNet model with the openmax layer decreases in most categories. This result suggests that while the openmax layer helps the model to detect noise images, it also sacrifices the performance of the model in classifying the map images in existing categories, with some map images being incorrectly classified as noise. We also

Table A1.: Classification accuracy of ResNet using softmax and openmax as the final output layer respectively.

	Softmax		Openmax		
	Correct	Incorrect	Correct	Incorrect	Noise
Africa	0.730	0.270	0.350	0.000	0.650
Antarctica	0.960	0.040	0.960	0.000	0.040
Asia	0.910	0.090	0.800	0.020	0.180
Europe	1.000	0.000	1.000	0.000	0.000
Global	0.980	0.020	0.880	0.000	0.120
North America	0.460	0.540	0.400	0.050	0.550
Oceania	0.150	0.750	0.100	0.050	0.850
South America	0.930	0.070	0.520	0.020	0.460
Noise	0.000	1.000	0.650	0.350	—

experimented with AlexNet and Inception-v3 by replacing their softmax layer with an openmax layer and obtained similar results. In sum, replacing the softmax layer with an openmax layer enables a CNN model to detect noise images that do not belong to any existing category. However, this new capability comes with a cost of decreased accuracy in correctly classifying the map images in existing categories.