

Transfer learning with convolutional neural networks for hydrological streamline delineation

Nattapon Jaroenchai^{a,b}, Shaowen Wang^{a,b,*}, Lawrence V. Stanislawski^c, Ethan Shavers^c, Zhe Jiang^d, Vasit Sagan^e, E. Lynn Usery^c

^a Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA

^b CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Urbana, IL, USA

^c U.S. Geological Survey, Center of Excellence for Geospatial Information Science, Rolla, MO, USA

^d Department of Computer & Information Science and Engineering, University of Florida, FL, USA

^e Department of Earth and Atmospheric Sciences, Saint Louis University, MO, USA

ARTICLE INFO

Handling Editor: Daniel P. Ames

Keywords:

Convolutional neural network

Deep learning

Remote sensing

Streamline analysis

Transfer learning

ABSTRACT

Hydrological streamline delineation is critical for effective environmental management, influencing agriculture sustainability, river dynamics, watershed planning, and more. This study develops a novel approach to combining transfer learning with convolutional neural networks that capitalize on image-based pre-trained models to improve the accuracy and transferability of streamline delineation. We evaluate the performance of eleven image-based pre-trained models and a baseline model using datasets from Rowan County, North Carolina, and Covington River, Virginia in the USA. Our results demonstrate that when models are adapted to a new area, the fine-tuned ImageNet pre-trained model exhibits superior predictive accuracy, markedly higher than the models trained from scratch or those only fine-tuned on the same area. Moreover, the pre-trained model achieves better smoothness and connectivity between classified streamline channels. These findings underline the effectiveness of transfer learning in enhancing the delineation of hydrological streamlines across varied geographies, offering a scalable solution for accurate and efficient environmental modelling.

1. Introduction

The National Hydrography Dataset (NHD) is a digital database of surface water features for the United States developed and managed by the U.S. Geological Survey (USGS) and partner organizations (Simley and Carswell, 2009). The NHD High Resolution (NHD HR) is the most current version of the NHD and is mapped at a scale of 1:24,000 or larger, except in Alaska where a scale of 1:63,360 or larger is available. However, because of the long-term data collection and compilation history of the NHD, some NHD data is from early twentieth-century topographic maps, and hence is of relatively lower quality with outdated content that compromised the usefulness of the data. The disagreement between the NHD and contemporary hydrologic and geomorphic conditions leads to erroneous reference data which can compound errors and/or result in inaccurate forecasting and interpretation of surface water processes (Jiang et al., 2022). In 2017, Stanislawski et al. (2017) leveraged parallel computing with various open-source geospatial analytical tools to automate the delineation of

elevation-derived streamlines for the conterminous United States. Despite this significant progress, extracted streamlines often need further manual editing, which is a labour-intensive and time-consuming process. Moreover, elevation-derived streamlines may be broken or obstructed by abrupt changes in elevation caused by roads, bridges, and valleys, which affect the performance of flow accumulation algorithms (Poppenga et al., 2013).

A variety of approaches have been developed and tested over time to obtain accurate and fine-scale digital hydrographic streamlines. Historic methods involve digitizing stream channels that are visible in aerial pictures and comparing them to the actual stream network (Guptill, 1983; Marks et al., 1984; Usery, 2011). With increasing computing power, digital methods for creating mapped hydrography can be employed using digital elevation models (DEMs). The basis of modelling stream networks using a DEM is that water flows in response to gradients in gravitational potential energy, accumulating to form channelized streams. This basis provides an alternate, more automated method of delineating flow channel networks that can be accomplished using

* Corresponding author. Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

E-mail address: shaowen@illinois.edu (S. Wang).

<https://doi.org/10.1016/j.envsoft.2024.106165>

Received 9 June 2024; Received in revised form 21 July 2024; Accepted 23 July 2024

Available online 24 July 2024

1364-8152/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

GIS-based flow accumulation algorithms.

Hydrologic modelling using flow routing methods based on gridded DEMs is the prevailing approach for large-scale drainage network extraction due to its simplicity, computational efficiency, and continuity in hydrological representation (O'Callaghan and Mark, 1984; Tarboton, 1997; Zhang et al., 2021). These methods typically involve steps such as pit filling, flow direction computation, and calculating the contributing area draining to each grid cell, along with defining a minimum threshold for contributing area to form linear drainage features (Stanislawski et al., 2017), as used broadly in USGS NHDPlusHR hydrologic data generation. Widely used software tools like Esri ArcGIS® Spatial Analyst Tools and the System for Automated Geoscientific Analysis (SAGA) facilitate the extraction of surface water drainage networks from DEMs using various flow routing methods such as D8 and D-infinity (Tarboton, 1997). Despite their widespread application, these methods have several limitations. In low-relief areas, flow direction models are less effective because they often restrict water flow to one or more adjacent pixels, leading to potential inaccuracies in flow path prediction. Moreover, the delineation process requires a unique threshold that often fails to accurately reproduce the actual network (Li et al., 2020; Passalacqua et al., 2010). Classic flow routing approaches also exhibit variable reliability across different drainage basins and grid cell sizes and shapes (Li et al., 2022), generally tending to overestimate the network and failing to predict channel heads reliably (Orlandini et al., 2011). As a result, manual edits are frequently needed to refine the flowlines produced by these methods (Lang et al., 2012).

There have been several studies on the use of a DEM to model stream networks (Ehlschlaeger, 1989; Metz et al., 2011; Tarboton and Ames, 2001; Wilson et al., 2008), as well as various enhancements to multiple techniques. Broadly, two methods exist to improve drainage system extraction using DEMs: 1) using high-resolution DEM; 2) improving the flow accumulation techniques. High-resolution DEMs derived from Lidar (Light Detection and Ranging) can provide details to detect and locate microtopographic features at the sub-meter scale. Hopkinson et al. (2009) found that when using three different data sources (Lidar, photogrammetry, and publicly available digital contour data) at two different resolutions (5 m and 25 m), a watershed area derived from digital contour data was overestimated by 15 percent when compared to the Lidar DEM. A similar result can be found in the study by Cao et al. (2021), which showed that a stream network modelled from Lidar DEMs is the most accurate among all types of DEMs due to the high point density, accuracy, and resolution. Lyu et al. (2021) proposed a method that extracts the vector of a drainage network directly from Lidar point clouds without limiting the number of water flow directions. This method helps to eliminate the limitation of imprecise DEMs. On the other hand, there are studies that attempt to improve stream network extraction by improving flow accumulation techniques. In a study by Zhang et al. (2021), a flow accumulation threshold (FAT) estimation method was developed using a statistical method to calculate the FAT from 47 other variables. A significant step in flow accumulation is depression filling, which helps correct DEMs to avoid pits and allows the simulated flow to move across the surface represented by DEMs. This filling or breaching algorithm has been improved by Barnes et al.'s (2021) study which runs 90–2600 times faster than the commonly used Jacobi iteration and produces more accurate output. Yet, it has been indicated that pit filling and breaching processes degrade the accuracy of modelled flow discharge (Rajib et al., 2020) and thus can negatively impact related flow-accumulation networks. Wang et al. (2021) proposed a fully automated drainage network extraction using Sentinel-2 satellite imagery. Yet, this strategy is ineffectual where surface water is occluded by vegetation or built structures. Even though modelling stream networks using DEMs has been significantly improved, it still entails several challenges that require expert knowledge and labour-intensive techniques in several steps, such as parameterizing extraction thresholds, eliminating flow obstructions, identifying headwater and sink locations, interpreting image data for validation, and

interactive editing of channels.

Recent advancements in deep learning have been widely acknowledged and adopted for various challenging feature recognition and detection tasks (LeCun et al., 2015; Schmidhuber, 2015; Kampffmeyer et al., 2016; Maggiori et al., 2017; Zhu et al., 2017; Xu et al., 2018). Convolutional Neural Networks (CNNs) and other deep learning algorithms have demonstrated remarkable performance in automatic object detection and recognition using close-range images for numerous computer vision applications. Their ability to learn directly from input data and eliminate human-involved processes has shown promising results for hydrography feature extraction applications. For instance, Li et al. (2021) utilized a U-net model, PA-UNet, to construct a robust waterbody extraction network from Sentinel-1 and Sentinel-2 imagery, achieving more than 95% extraction accuracy (based on mapped waterbodies). Similarly, Liu et al. (2020) proposed LaeNET to automatically extract lake areas and shorelines from Landsat-8 imagery, achieving an F1-score of 99.41%. Studies by Chen et al. (2020, 2018) also demonstrated the superiority of deep learning methods over traditional methods for extracting bodies of water from high-resolution remote sensing imagery.

However, while these CNN and deep learning models provide automated hydrography extraction for waterbodies, their application for precise, efficient, and fine-scale delineation of stream networks has been limited. Xu et al. (2021) were among the first to address this problem using a CNN for hydrographic streamline extraction, proposing an attention U-net model that achieved F1-scores averaging 11.2% better than baseline machine learning approaches and provided improved smoothness and connectivity across identified streamline channels. Lin et al. (2021) also used a machine learning approach to estimate river network drainage density based on global watershed-level climatic, topographic, hydrologic, and geologic conditions, showing better agreement with Landsat-derived river center lines than methods proposed by Verdin (2017) and Grill et al. (2019). Additionally, Stanislawski et al. (2021) tested the extensibility of Xu et al.'s (2021) model by training it on a minimal portion of a study area and using the prediction results to constrain the D8 flow accumulation method for streamline extraction across a larger region.

Despite these advancements, CNN and deep learning approaches for DEM-based hydrography feature extraction face significant challenges. These methods require large datasets to train effectively and leverage their ability to encode fine details, which entails substantial training time and computational power dependent on the model architecture and the number of trainable parameters (Stanislawski et al., 2021). Furthermore, as emphasized by Shen et al. (2021), most models are trained locally with samples from the site of interest, resulting in poor performance outside the training region and insufficient testing of potential prediction failures. These challenges underscore the need for adaptation of transfer learning strategies, which can leverage knowledge from one domain to another, potentially reducing the time and data required for training and enhancing model performance in a variety of geographic areas.

Transfer learning is a branch of machine learning that leverages knowledge from one domain to another domain. It serves as an alternative to building models from scratch as in traditional machine learning (Gao et al., 2008). Transfer learning is based on the premise that knowledge gained from training a neural network on one task can be used to improve the training process of a network on a new but related task. Transfer learning has become an active topic in the context of deep learning. Transfer learning, when done properly, can reduce the time and quantity of data required for training, resulting in a shorter, less computationally intensive, and hence less costly training process. Yosinski et al. (2014) have found that initialization using transferred information improves generalization performance. However, it is important to recognize its limitations. One such limitation is negative transfer (Wang et al., 2019), which occurs when the knowledge acquired from the source task negatively impacts performance on the target task. To better avoid negative transfer, one can carefully select the source task

and to ensure that it is sufficiently different from the target task. Additionally, appropriate fine-tuning techniques can help mitigate the effects of negative transfer.

The U-Net architecture was originally proposed by [Ronneberger et al. \(2015\)](#) for medical applications and is widely used for image segmentation tasks ([Xiao et al., 2023](#); [Aggarwal et al., 2022](#); [Colaco et al., 2022](#); [Gowroju et al., 2022](#)). The architecture is comprised of two parts: i) the encoder, and ii) decoder. The encoder decomposes an image into numerous low-level feature map layers. The corresponding decoder layers use up-sampling layers to process the features back to the input size. The encoder CNNs in this study are constructed utilizing the networks pre-trained on an image classification task using the ImageNet dataset ([Deng et al., 2009](#)). Previous studies have shown that utilizing CNNs pre-trained on ImageNet as the encoder instead of retraining the entire network from scratch results in improved performance for many tasks ([Mukhlif et al., 2023](#); [Ünal and Aktaş, 2023](#); [Li et al., 2022](#); [Wang et al., 2022](#)). In this paper, we explore whether this approach can be applied to the streamline delineation task and yield similar benefits.

We describe and test a transfer learning approach based on fine-tuning deep convolutional neural networks to overcome challenges of streamline delineation across spatially heterogeneous geographic areas. We evaluate and compare the recently proposed, attention U-net model developed by [Xu et al. \(2021\)](#), with neural networks pre-trained on the ImageNet dataset using three accuracy metrics: F1 score, recall, and precision, with the focus on F1 score. We also determine the degree to which such an approach can be scaled to large geographic areas.

Thus, we contribute to this growing body of knowledge in two ways via this paper: (1) comparing the effectiveness of knowledge transfer for U-net models with encoders trained on general data and the attention U-net model ([Xu et al., 2021](#)) built specifically for the streamline delineation task, and (2) demonstrating the models' effectiveness when transferred to different geographies.

2. Methods

In this section, we first describe the study areas and data preparation. Then, we give detailed descriptions of the training methods used in the experiments. Finally, we explain the model evaluation methods.

2.1. Study areas and data preparation

We used two locations to investigate the transferability of the models: i) a small watershed in Rowan County in North Carolina, and ii) a Covington River watershed area in Virginia.

2.1.1. Rowan County, North Carolina

The data for a small watershed in Rowan County, North Carolina ([Fig. 1](#)) are retrieved from the study by [Xu et al. \(2021\)](#). This area is 6.3 square kilometers (km^2) in a mountainous rural section of west-central North Carolina that is dominated by forested land cover type. The area encompasses a set of tributaries that are partially covered by tree canopy and that flow into Second Creek, which is a primary stream network feature of the 12-digit NHD Hydrologic Unit (HU12) 030401020504 watershed. The Rowan area is in the Central Interior and Appalachian Ecological Division ([Comer et al., 2003](#)). The dataset contains 1400 training samples and 30 validation samples extracted from the northern of the area. The test data is the entire southern of the area. A set of intermittent stream head data, verified in the field and gathered between 2013 and 2014, was enhanced with on-screen editing to produce reference information ([Shavers and Stanislawski, 2018](#)). A 3-m buffer was created around the reference streamline to mimic the breadth of stream channels.

Eight raster layers are stacked together to create the dataset. These raster layers are derived from Quality Level 1 (QL1) ([Heidemann, 2012](#)) Geiger-mode Lidar point cloud data. The eight 1 meter spatial resolution layers include: (1) a DEM derived from the lidar ground return points; (2) geometric curvature determined from the DEM; (3) a topographic position index (TPI) derived from the DEM using a 3-cell by 3-cell window; (4) a TPI derived from the DEM using a 21-cell by 21-cell window; (5) zenith angle positive openness derived from the DEM using a 10-cell radius with 32 directions ([Doneus, 2013](#)); (6) return intensity determined from the lidar ground points averaged with inverse distance weighting using 10 nearest points; (7) point density for return points between zero and 1 foot above ground; and (8) point density for return points between 0 and 3 feet above ground. The statistics of each raster layer are presented in [Table 1](#).

2.1.2. Covington River watershed, Virginia

The other study area for this research is the 12-digit USGS NHD

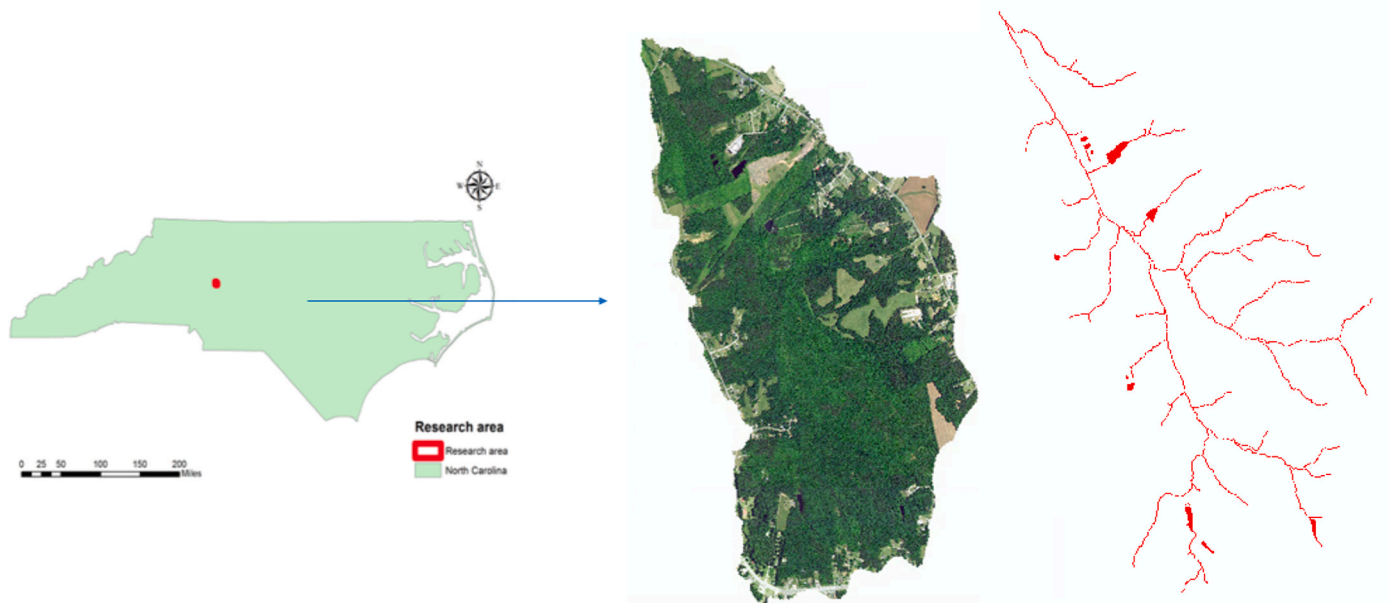


Fig. 1. Rowan County area (left: location of study area within North Carolina; middle: a 1-m resolution image of the study area from the National Agriculture Imagery Program (NAIP); right: reference stream features). Adapted from [Xu et al. \(2021\)](#).

Table 1

Summary statistics raster images generated for the study area in Rowan County, North Carolina.

Raster image name	Minimum	Maximum	Mean	Standard Deviation	Range
Digital elevation model (meters)	194.11	256.19	229.07	12.96	62.07
Geometric curvature	−97.25	97.93	0.01	3.05	195.18
Topographic position index (3 × 3 window)	−8.59	5.58	6.38	0.18	14.167
Topographic position index (21 × 21 window)	−13.62	13.29	0	0.93	26.91
Openness (R10, D32) degrees	21.52	118.8	83.41	7.35	97.28
Return intensity	0	55185.39	29047.18	10624.11	55185.39
Return point density 1 ft above ground (points per m ²)	0	0.94	0.02	0.04	0.94
Return point density 3 ft above ground (points per m ²)	0	2.89	0.12	0.23	2.89

Note. Adapted from Xu et al. (2021).

Hydrologic Unit (HU12) 020801030302 watershed that includes primary tributaries of the Covington and Rush Rivers. This watershed is in Rappahannock County, northern Virginia (Fig. 2) and the study area is 108 km². The Blue Ridge Mountain range influences the climate in the area to the west and the Atlantic Ocean and the Chesapeake Bay to the east. The temperature ranges from 26 °F (−3 °C) to 90 °F (32 °C). The two dominant land cover types are forest and agricultural land. The stream channels are occluded by the canopy in the forested areas and exposed in the agricultural areas. The elevation ranges from 125 to 1040 m (m) above sea level. The watershed is in the Central Interior and Appalachian Ecological Division (Comer et al., 2003).

The NHD flowline and the NHD waterbody feature datasets are used

for reference data. The reference NHD features were collected through a USGS contract in 2017 and were compiled to meet 1:24,000 or larger scale representations. Contractors used best available methods to extract hydrographic features from Lidar point cloud and derived elevation models. Lidar point cloud data were collected during leaf-off season in 2014–2015 to meet USGS Quality Level 2 (QL2) (Heidemann, 2012). Extracted hydrographic features were manually edited and verified as needed based on an interpretation of high-resolution National Agriculture Image Program (NAIP) orthoimage data. The collected NHD flowlines, polygonal lakes, reservoirs, and streams were rasterized to 1 meter resolution, with the flowlines buffered 3-m on each side, which simulates a 6-m minimum channel width for all flow network features. Wider raster streams are represented where polygonal streams are included in the reference data.

Eight co-registered 1 meter resolution Lidar and elevation-derived raster data layers were used for training, validation, and testing in the research. The 1 meter raster layers include: (1) DEM; (2) geometric curvature determined from the DEM (Sangireddy et al., 2016); (3) slope derived from the DEM; (4) mean nadir angle positive openness derived from the DEM using a 10-cell radius with 32 directions (Doneus, 2013); (5) topographic position index (TPI) derived from the DEM using a 21-cell by 21-cell window; (6) return intensity determined from the Lidar ground points averaged with inverse distance weighting using 10 nearest points; (7) Geomorphons type estimated as one of ten common landforms derived from the DEM using a 10-cell radius (Jasiewicz and Stepinski, 2013); and (8) TPI derived from the DEM using a 3-cell by 3-cell window.

The eight raster layers are presented in Fig. 3 with the summary statistics presented in Table 2. To create training, validation, and test datasets, the raster layers (1) to (8) pixel values are normalized to the scale from 0 to 255 using equation (1).

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} * 255 \quad (1)$$

Normalization helps avoid numerical instability caused by differences in the range (presented in Table 2) by changing the values to use a common scale, which can improve convergence speed and reliability for neural network training. Then, all sample patches are extracted with the size of 244 pixels by 244 pixels, and contain the eight raster layers (1) to (8) at 1 meter resolution (Fig. 4).

To construct the training dataset for the Covington watershed, an initial 200 sample patches are extracted. Then, we used image augmentation to randomly rotate each sample patch by 30°–150° and 210°–330°, randomly rescaled each sample by the factor of 0.5–0.8 and

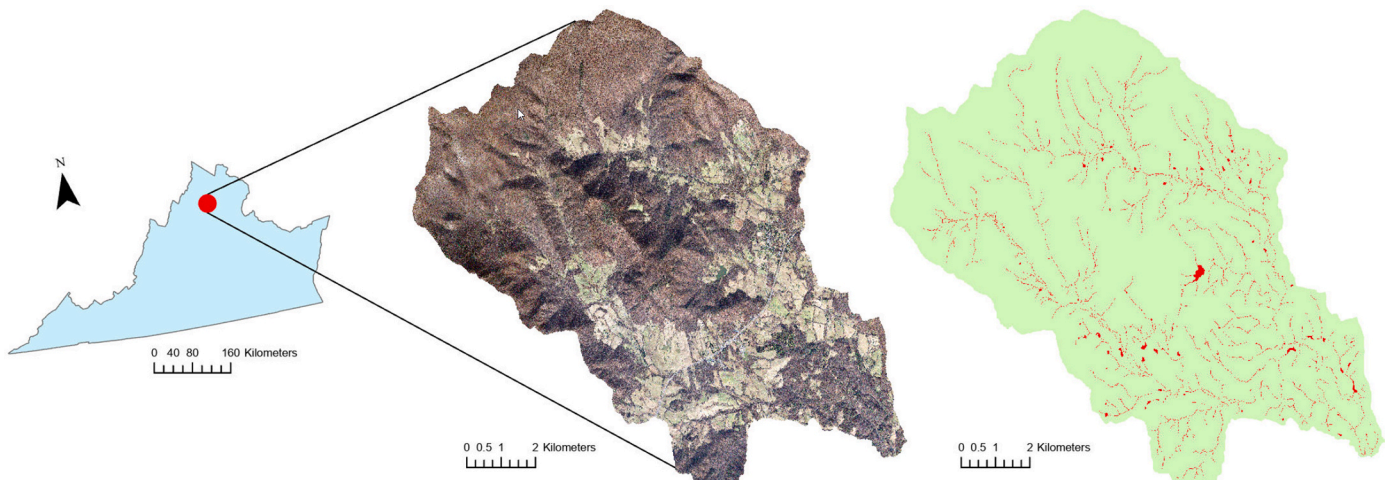


Fig. 2. Covington area (left: Virginia, USA, showing the location of the study area as a red circle; middle: a 1 meter resolution image of the study area from National Agriculture Imagery Program (NAIP); right: reference stream feature shown in red pixels on light green background).

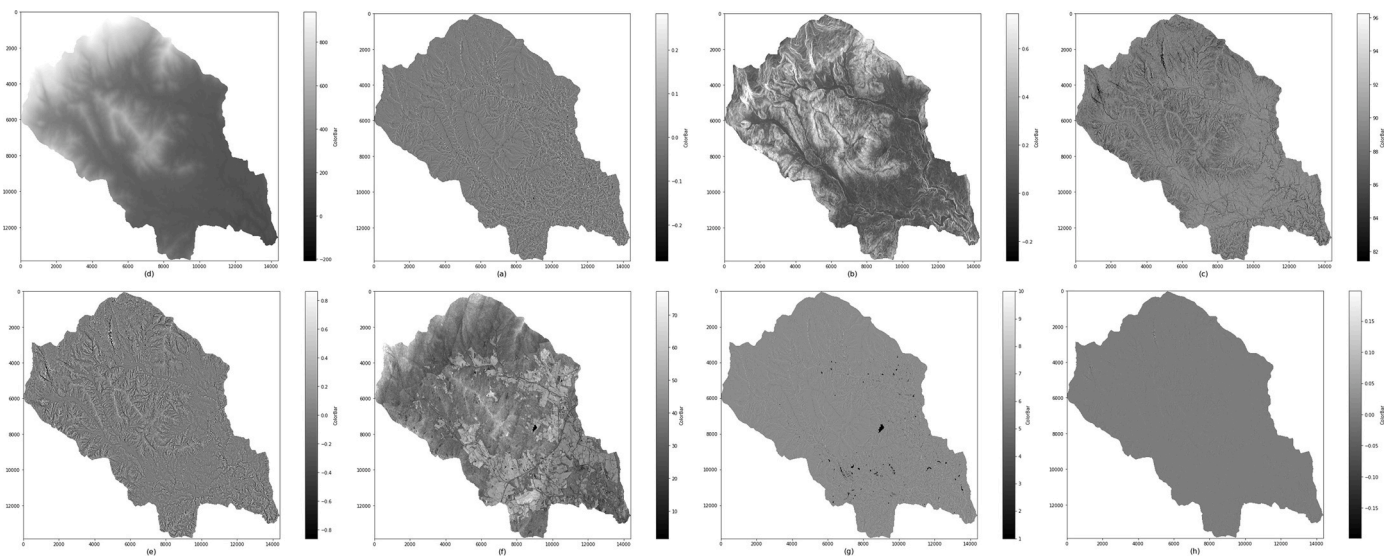


Fig. 3. 1 meter resolution raster images: (a) Digital Elevation Model (DEM); (b) Geometric curvature derived from DEM; (c) Slope; (d) Positive openness; (e) TPI with moving window size 21; (f) Lidar reflectance intensity; (g) geomorphon 10 most common landform types; (h) TPI with moving window size 3.

Table 2

Summary statistics for 1 meter resolution raster images generated for the Covington River watershed, VA.

Raster image name	Minimum	Maximum	Mean	Standard Deviation	Range
Digital elevation model (DEM) (m)	125.4523	1039.1520	365.8976	190.9829	913.6997
Geometric curvature	-1.9900	1.9974	0.0001	0.0941	3.9957
TPI with moving window size 3	-5.7020	5.7213	0.00000235	0.0661	11.4232
TPI with moving window size 21	-14.6981	12.8009	0.0001	0.2873	27.4990
Positive openness	45.3490	162.6082	88.8340	2.4694	117.2592
Lidar reflectance intensity	0.0000	255.0000	39.3197	12.6086	255.0000
Slope data (degree)	0.0000	14.2646	0.2323	0.1711	14.2646

Note. Geomorphons is an integer-coded discrete class therefore we do not include the statistics in this table.

1.5–2.0, sheared each sample randomly between -30° and 30° , and mirror each sample horizontally, resulting in six augmented samples per patch for a total of 1400 samples for the training dataset. For the validation dataset, 200 unaugmented sample patches are extracted. The training and validation dataset patches were extracted from the northern region of the study area. The entire southern region of the study area was used as the test dataset to evaluate the prediction accuracy and the generalizability of the models. Fig. 5 depicts the northern and southern regions used for the Covington datasets (i.e., training + validation and testing dataset respectively).

2.2. Model creation and training

The baseline model used for this study is the attention U-net as modelled by Xu et al. (2021), which were trained on the data layers described in section 3.1.1 and Table 1. All other models were created using a GitHub library called Segmentation Models (Yakubovskiy, 2019), a Python package based on Keras (Chollet and others, 2015) and TensorFlow (Martín and others, 2015); this library includes four popular segmentation architectures and ImageNet pre-trained backbones that can be merged into the segmentation model architecture of choice. The library supports the following architectures: (i) U-net (Ronneberger et al., 2015), (ii) FPN (Kirillov et al., 2017), (iii) LinkNet (Chaurasia and Culurciello, 2017), and (iv) Pyramid Scene Parsing Network (PSPNet) (Zhao et al., 2016). In this study, we chose U-net as the model architecture because it has been found to be effective for creating segmentation masks (Siddique et al., 2021), it achieves high accuracy for solving image segmentation problems (Ronneberger et al., 2015), and it was successfully used previously by Xu et al. (2021) on our baseline model dataset.

The U-net architecture is created in an encoder-decoder form using skip connections, allowing the use of a variety of ImageNet pre-trained backbones as the encoder (i.e., VGG16 as shown in Fig. 6) to evaluate their capability for streamline delineation. The pre-trained backbones include the following: ResNet34 (He et al., 2015), ResNet50 (He et al., 2015), ResNet101 (He et al., 2015), ResNet152 (He et al., 2015), DenseNet121 (Huang et al., 2016), DenseNet169 (Huang et al., 2016), DenseNet201 (Huang et al., 2016), VGG16 (Simonyan and Zisserman, 2014), VGG19 (Simonyan and Zisserman, 2014), InceptionResNetV2 (Szegedy et al., 2016), and Inceptionv3 (Szegedy et al., 2015). These backbones demonstrate significant gains over previous state-of-the-art methods and are well recognized in the literature. In accordance with the documentation of the Segmentation Models (Yakubovskiy, 2022), an extra convolutional layer was added to map the eight input channels of the input to the three input channels of the ImageNet pre-trained backbones (refer to the left most layer in Fig. 6 as an example).

The models created with 'Segmentation Models' were compiled using the "Adam" (Zhang, 2018) optimization algorithm while a custom "dice_coef_loss" loss function and a custom defined "dice_coef" function were generated for evaluation. The dice coefficient loss function is specifically designed to address the imbalanced nature of the streamline delineation task. In this context, the dice coefficient, also known as the Sørensen–Dice index, is a measure of overlap between two samples. It ranges from 0 to 1, where 1 indicates perfect and hence complete overlap. The dice loss is defined as 1 minus the dice coefficient, making it a suitable choice for optimizing models in segmentation tasks where the positive class (stream pixels) is significantly outnumbered by the negative class (non-stream pixels). This custom loss function helps the model to focus more on the underrepresented class, thereby improving the accuracy of stream detection and overall model performance. By

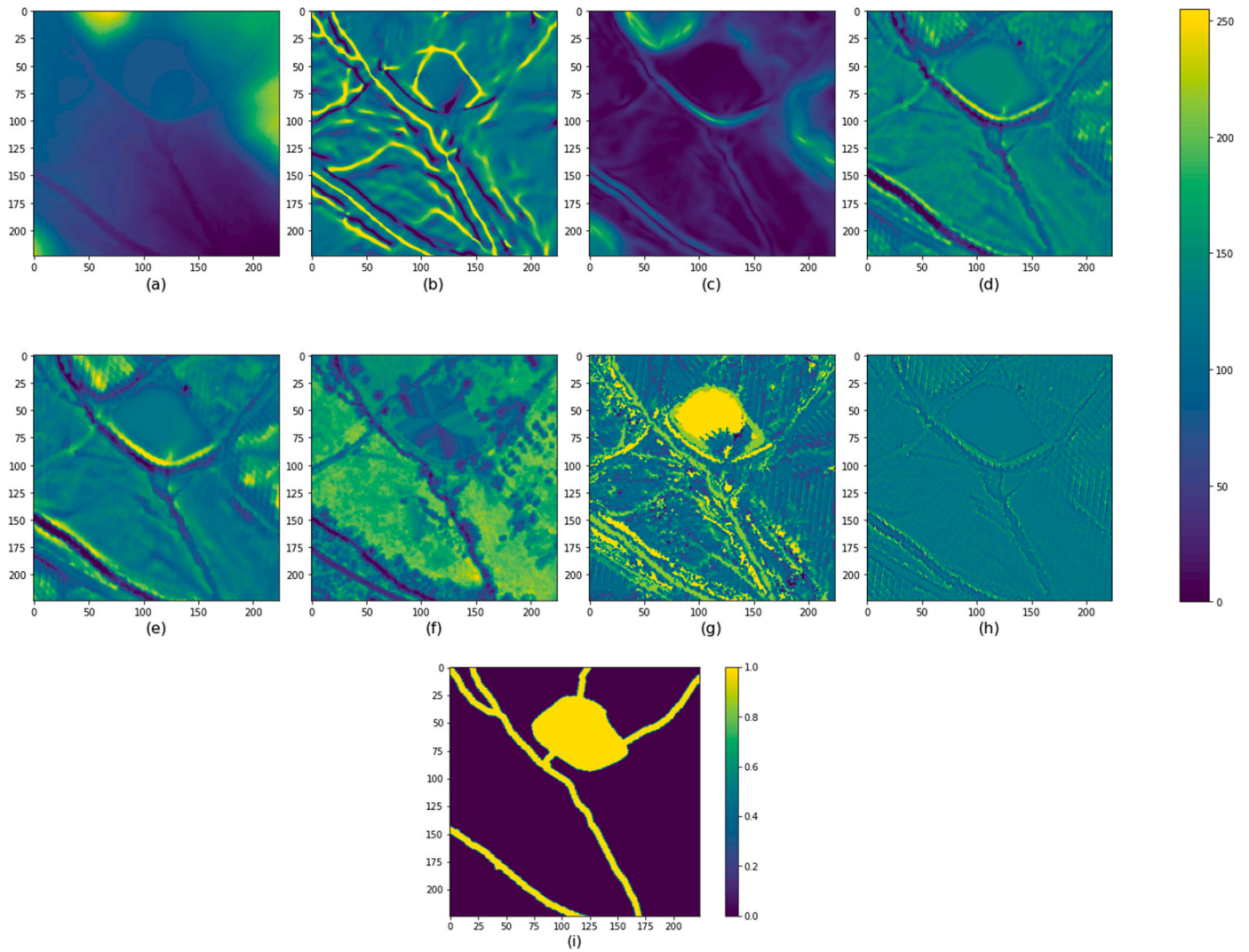


Fig. 4. An example of sample patches: (a) Digital Elevation Model (DEM); (b) Geometric curvature; (c) Slope; (d) Positive openness; (e) TPI with window size 21; (f) Lidar reflectance; (g) Geomorphons; and (h) TPI with window size 3; (i) Reference label where yellow represents water and purple represents non-water. Axes values are distance in meters. The color bar values are the normalized raster values.

minimizing the dice coefficient loss, the model can better capture fine details and maintain the continuity of stream networks, which is critical for high-resolution hydrological applications.

Training was initiated with the Keras fit function. In addition to controlling the training process, we used callback objects. We applied the following modules in the callbacks: “ReduceLROnPlateau”, “EarlyStopping”, and “ModelCheckpoint”. The “ReduceLROnPlateau” is used to decrease the learning rate when the monitored metric has stopped improving. In our implementation, the callback monitors the validation loss value (“val_loss”). The learning rate will decrease by the factor of 0.7 if no decreasing of the “val_loss” is seen for ten epochs. The minimum learning rate was set to 10^{-9} . “EarlyStopping” callback is used to stop the training process if the monitored metric stops improving for some number of epochs. In our case, the validation loss value was monitored by the callback, and the training process was set to stop after ten epochs of no improvement. Finally, “ModelCheckpoint” was used to save the current best model in terms of loss value on the validation set.

2.3. Model training and evaluation

We used two approaches to train the models in this study: training from scratch and fine-tuning for transfer learning. In the training from scratch strategy (TS), we randomly initialized the weights of the models

and trained the models from scratch. The F1-scores of the attention U-net model are used as the baseline to evaluate and compare the performance of the other U-net models. For the fine-tuning strategies, there were two levels of the fine-tuning process: (i) transfer to different task (TL1) and (ii) transfer to different geography (TL2).

For TL1, we use the weights of ImageNet backbones and fine-tune it to the streamline delineation task on the Rowan County dataset. Our goal was to investigate whether fine-tuning a pre-trained ImageNet model could achieve better performance than directly training an attention U-Net from scratch. The training process of TL1 was done in two steps. In the initial step of the training process, the encoder part of a network was initialized with the ImageNet pre-trained weights and then frozen. Then, the network was trained with the initial value of 10^{-3} for the learning rate, which is the default value of the optimizer function. In the second step of the training, the whole network was unfrozen and trained with a significantly smaller learning rate, 10^{-5} . The lower learning rate in the second phase was used to preserve the knowledge learned in the first phase and fine-tune the network to the input dataset. In the fine-tuning process of TL2, our goal was to compare the performance of different models when being transferred to a new geographic area. Specifically, we initialized a model with the weights trained on Rowan County dataset (e.g., those models from TL1), then, fine-tuned the model with the Covington area dataset. The network was trained

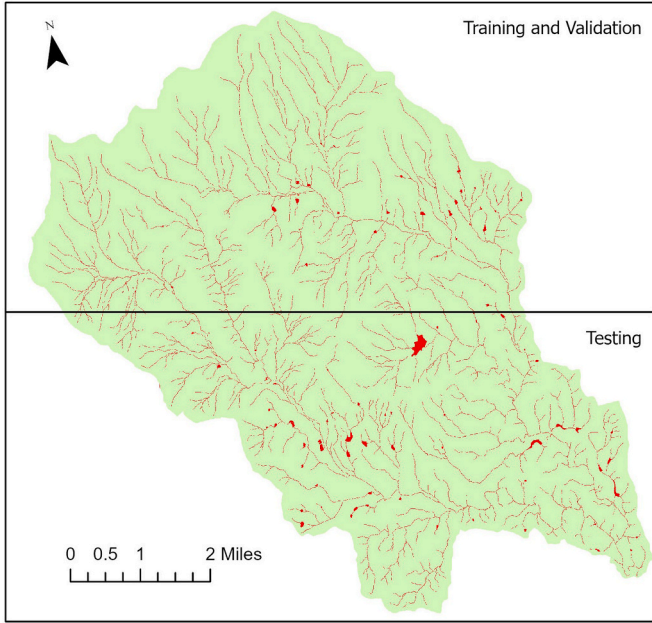


Fig. 5. Visualization of the northern and southern regions of Covington, which are used to generate the training dataset and testing dataset, respectively.

in two steps similar to the TL1.

The three metric functions precision (3), recall (4), and F1 score (5) were used to evaluate the performance of the models with test datasets. The three matrices are defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{F1 Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

where TP is the number of true positive (correctly predicted water) pixels,

FP is the number of false positive (incorrectly predicted water) pixels, and.

FN is the number of false negative (incorrectly predicted non-water) pixels.

2.4. Hardware and software

All experiments were performed using NVIDIA GeForce RTX 2080 SUPER and Tesla T4 on Google Colab (Bisong, 2019). We used Python 3.7.9 and Keras Python Library (Chollet and others, 2015) version 2.3.1 with the backend of TensorFlow (Martín and others, 2015) version 2.1.0 for model construction. We also used Python libraries including scikit-image 0.18.1, scikit-learn 0.24.0, Rtree, 0.9.4, GDAL 3.0.2, and NumPy 1.19.4. The software environment is supported by Anaconda 4.10.0.

3. Results and discussion

We designed two experiments. In the first experiment, we compared the performance of the U-net architecture with different ImageNet pre-trained backbones (TL1) with the baseline attention U-net model (TS) using the data collected from the Rowan County study area (Xu et al., 2021). The test dataset from the same area were used to calculate precision, recall and F1-score to compare the performance of the models. In the second experiment, we examined the transferability of the fine-tuning models across different geographic areas (TL2). The top three ImageNet pre-trained models from the first experiment and the attention U-net model were fine-tuned with the data collected from the Covington study area described in section 3.1.

In the first experiment, we compared the performance of models pre-trained on a generic image dataset with the U-net model trained from

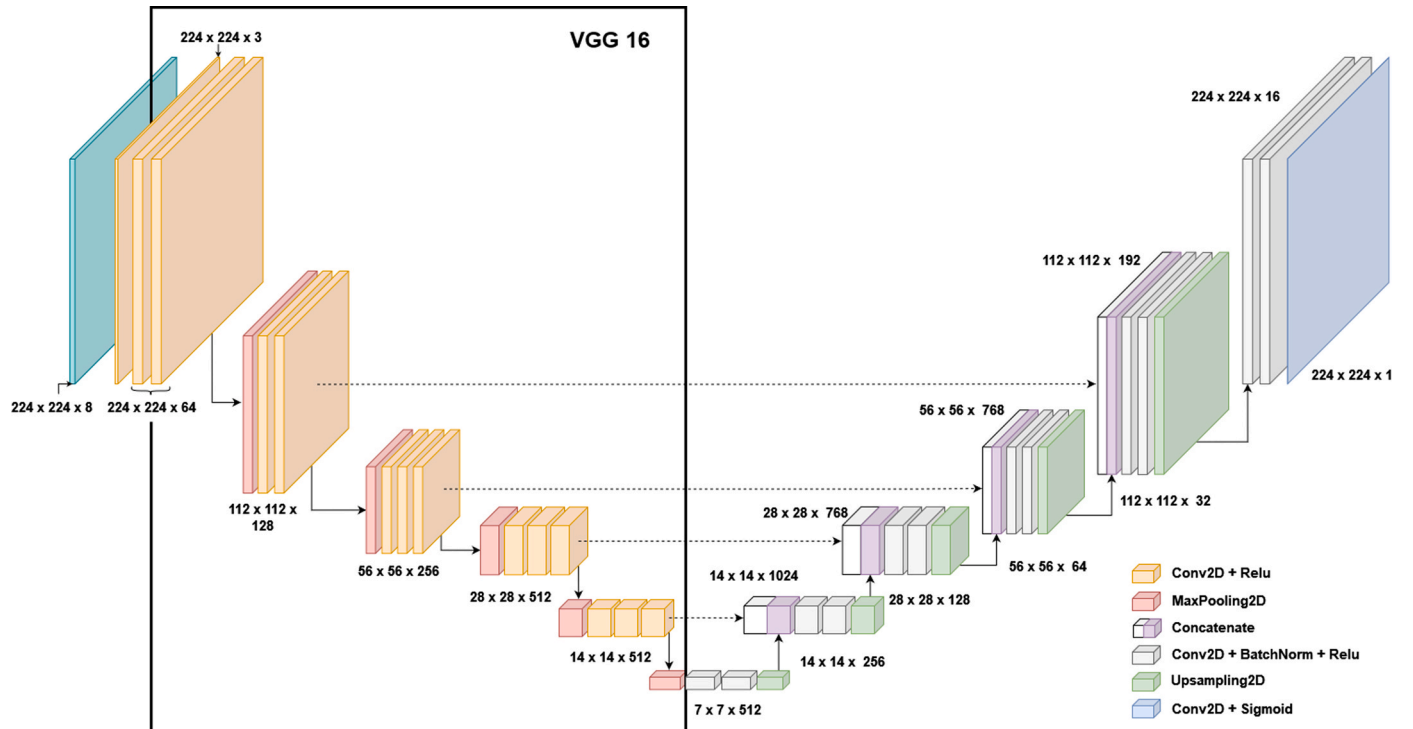


Fig. 6. The U-net architecture with the VGG16 backbones, one of the ImageNet pre-trained used in the study.

scratch on the dataset specific to the task. For this purpose, we used eleven models with different ImageNet pre-trained backbones and applied the TL1 strategy with the training dataset from the Rowan County area. In comparison, we applied the TS strategy to the eleven models. Additionally, we examined the attention U-net model using the same dataset. It is worth noting that the performance of the attention U-net model reproduces scenario 1 from Xu's (2021) study. In Xu's (2021) study, the top part of the study area was utilized as training samples, while the bottom part served as the testing dataset (Xu et al., 2021). The performance of each model is presented in the order of F1-score of stream class in Table 3. From the data, it is evident that models initialized with the ImageNet pre-trained weights outperform those initialized with random weights. On average, there are an increase of 4.67% in F1-score, with the InceptionResNetV2 model showing the highest improvement at 6.97% and DenseNet121 at the lower end with an improvement of 1.88%. These results underscore the benefits of transfer learning for image classification tasks. In particular, the knowledge derived from ImageNet offers significant advantages to streamline delineation tasks. Further, our data analysis suggests that this enhanced performance is chiefly attributed to the use of ImageNet-initialized weights and not necessarily due to differences in the model architectures themselves.

When we examined the results among the models initialized with ImageNet pre-trained weights in the second column of Table 3, we found that the best model has an F1-score that is 6.15% higher than the worst models, DenseNet169 and VGG19, respectively. It was observed that there was slightly better performance by smaller models with fewer parameters, compared to the larger models with more parameters. The top three models averaged 21,434,778 parameters, whereas the bottom three models averaged 40,378,348 parameters. The baseline model performed slightly better than all variants of VGG and Inception backbones, with 2.32% improvement compared to the worst model, VGG19. All variants of DenseNet and ResNet backbones outperformed the baseline model by as much as 4%, compared to DenseNet169. The unique features and strengths of these architectures, such as ResNet's residual connections and DenseNet's dense connectivity, may have contributed to their superior performance in capturing the complex patterns and details required for accurate streamline delineation.

It is worth noting that not only did the fine-tuned models outperform the baseline model, but the decreasing time to convergence and the decreasing training time was also significant. The baseline model took 130 epochs, whereas the fine-tuned models took 40 to 70 epochs, and the training time per epoch was decreased by 38% on average between the baseline U-net model and the fine-tuned models. From the results of

this experiment, it is evident that transfer learning applied to the secondary task (streamline delineation), as opposed to the primary task (ImageNet classification), performs better than training the models with randomly initialized weights. This can be attributed to the hierarchical feature learning and the transferability of the learned features across different domains.

A crucial factor to consider regarding the study area's heterogeneity is the substantial disparity between the data derived from Rowan County and the Covington area, as evidenced by the t-SNE (t-distributed Stochastic Neighbour Embedding) and PCA (Principal Component Analysis) plots presented in Fig. 7. These plots illustrate a clear demarcation between the two regions, indicative of hydromorphic differences in the two study areas. Particularly in the PCA plot, not only are the data points from each area noticeably segregated, but the distribution of stream and non-stream pixels also exhibit discernible dissimilarities. The data points associated with the Covington area are more widely dispersed, indicating a greater diversity in both stream and non-stream pixels than the Rowan County data. It is important to note that the dissimilarity in data distributions is not solely due to geographical disparities but is also influenced by variations in input data layers between the Covington and Rowan County datasets. These differing input layers contribute to the observed segregation in the PCA plot, reflecting variations in environmental, hydrologic, geomorphic and geological factors between the two regions, further underscoring the complexity of factors influencing pixel distribution in our study areas.

In the following experiment, we investigated the generalizability of the models across diverse geographical contexts. Specifically, we assessed the effectiveness of the top-performing models in terms of F-1 score from the previous experiment, namely DenseNet169, ResNet50, DenseNet121, and the attention U-net model, in the context of the Covington study area. To accomplish this, we employed the transfer learning strategy TL2, wherein the models are fine-tuned using knowledge acquired from the Rowan County area while adapting to the specific characteristics of the Covington area. Hence, the Rowan County area serves as the source location, and the Covington area serves as the target location, allowing the preservation of acquired knowledge while fine-tuning for optimal performance in the Covington area.

To establish a performance benchmark for these chosen models, we applied the TL1 strategy to the three selected ImageNet pre-trained models, and the TS strategy to the attention U-net, using the Covington area training dataset. All models were then assessed using the Covington test dataset. As indicated in Table 4, the findings mirror those from the initial experiment outlined in Table 3. All three ImageNet pre-trained models surpassed the performance of the attention U-net model, with the highest performance gain seen in ResNet50 at 13.14%. Remarkably, despite fine-tuning the models or training them from scratch using the Covington dataset, and then assessing them with the same area's test dataset, the models demonstrate a significant reduction in effectiveness, with an average decrease of 30% across all models. This could be attributed to the quality of the reference hydrographic data in Covington and the superior quality (QL1) of Lidar data in Rowan County. However, the regional and landcover differences likely also have had an influence on these results.

We subsequently fine-tuned all four models utilizing the TL2 strategy. Models were initialized with the weights from the Rowan County models in the first experiment and fine-tuned using the Covington training dataset. The performance of all models was evaluated with the Covington test dataset. Results are presented in Table 5. In comparing the performance obtained from direct training on the Covington dataset, the fine-tuning of Rowan County models to accommodate the Covington dataset resulted in significant enhancements across all models. On average, there was a performance enhancement of 2.25%, and ResNet50 noted a maximum improvement of 3.99%. Moreover, this strategy yielded the highest performance for the Covington area, achieving an F1-score of 65.58% with ResNet50.

Observations indicated that the attention U-net model consistently

Table 3

Performance of the models trained or fine-tuned and evaluated on the Rowan County test dataset, ordered by the F1-score of models initialized with ImageNet pre-trained weights. Precision and recalls of these models are provided in Appendix A.

Models	F1-Score for Stream Class		#Parameters
	Initialized with ImageNet pre-trained weights	Initialized with random weights	
DenseNet 169	85.11%	81.80%	19,545,964
ResNet 50	83.77%	78.00%	32,587,253
DenseNet 121	83.73%	81.85%	12,171,116
DenseNet 201	82.95%	80.37%	26,404,716
ResNet 34	82.51%	76.68%	24,482,293
ResNet 101	81.78%	75.85%	51,631,605
ResNet 152	81.65%	78.40%	67,321,333
Attention U-net	81.28%		53,508,217
VGG16	80.98%	74.23%	23,778,412
InceptionResNet V2	80.76%	73.79%	62,087,692
Inceptionv3	80.42%	75.45%	29,959,244
VGG19	78.96%	74.79%	29,088,108



Fig. 7. t-SNE (left) and PCA (right) plots illustrating the substantial disparity between data derived from Rowan County and the Covington area. The PCA plot shows clear segregation of data points from each area, with Covington data points more widely dispersed, indicating greater diversity in stream and non-stream pixels compared to Rowan County. This segregation is influenced by both geographical disparities and variations in input data layers, reflecting different environmental and geological factors between the regions.

Table 4

Performance of the models when fine-tuned or trained from scratch on the Covington area and the Rowan County study area (Note that the Rowan County results are the same as Table 3). Precision and recalls are provided in Appendix B.

Models	F1-Score for Stream Class		#Parameters
	Covington	Rowan County	
ResNet50	61.59%	83.77%	32,587,253
DenseNet121	55.67%	83.73%	12,171,116
DenseNet169	51.25%	85.11%	19,545,964
Attention U-net	48.45%	81.28%	53,508,217

Table 5

Performance of models when initialized with the weights from the Rowan County area and fine-tuned further with the data from the Covington area. Precision and recalls are provided in Appendix C.

Models	F1-score	
	Rowan County Fine-tuned to Covington (TL2)	Baseline in Covington
ResNet50	65.58%	61.59%
DenseNet121	56.66%	55.67%
DenseNet169	54.77%	51.25%
Attention U-net	48.95%	48.45%

held the lowest rank across all training strategies, whether trained from scratch (TS) using the Covington area dataset or through the transfer from Rowan County to Covington area (TL2). The relatively underwhelming performance of the attention U-net model, compared to other ImageNet models, can potentially be attributed to the intricacies inherent in the ImageNet datasets. The datasets employed for training the attention U-net model comprise a total of 2800 samples, with the Rowan County and Covington area contributing 1400 samples each. In contrast, the models pre-trained on ImageNet utilized over 14 million images, characterised by their complexity and richness, attributable to the variations both within a single class and across different classes.

In the final stage of the study, we assessed the performance of a variety of deep learning models as well as traditional machine learning methods pertaining to the extraction of hydrographic features and streamline delineation within the Covington area. The results, presented in Table 6, reinforce the outstanding performance of the ResNet50

model, which was trained utilizing the TL2 strategy. Other deep learning models, including DenseNet169 and DenseNet121, also demonstrated robust performance, thereby underlining the merits of transfer learning and fine-tuning strategies. This is in contrast to the attention U-net model, which trailed behind all other deep learning models in achieving an F1-score of only 48.95%. In a general context, traditional methods such as Random Forest, Support Vector Machine, and Gradient Boosting Classifier displayed substandard performance, reflecting their restricted effectiveness in the task of streamline delineation. These findings accentuate the superiority of deep learning models, specifically those fine-tuned with data sourced from the target area, over traditional methods for the tasks of hydrographic feature extraction and streamline delineation. This emphasizes the critical role of transfer learning and fine-tuning strategies in accomplishing superior performance and adaptability across varied geographical contexts.

We visualized the prediction results of transfer learning models (ResNet50, DenseNet121, DenseNet169, and Attention U-net) alongside traditional machine learning methods (Random Forest, SVM, and GBC). In Fig. 8, the large-extent visualization shows that transfer learning models generate smoother, more connected streamlines compared to the fragmented channels produced by traditional methods, which also predominantly identify major channels with significant overestimation of stream pixels by GBC while failing to delineate smaller channels accurately. Further examination in Fig. 9 focuses on zoomed-in contexts, where ResNet50 is particularly noteworthy for producing cleaner and more well-connected channels, highlighting its superior performance in capturing detailed hydrographic features. These visualizations collectively underscore the enhanced capability of transfer learning models in accurately and efficiently delineating streamlines compared to traditional machine learning methods.

Table 6

Comparison of the F1 scores between deep learning models and the traditional methods. Precision and recalls are provided in Appendix D.

Models/methods	F1-score in Covington
ResNet50	65.58%
DenseNet121	56.66%
DenseNet169	54.77%
Attention U-net	48.95%
Random Forest	46.14%
Support Vector Machine	39.81%
Gradient Boosting Classifier	30.09%

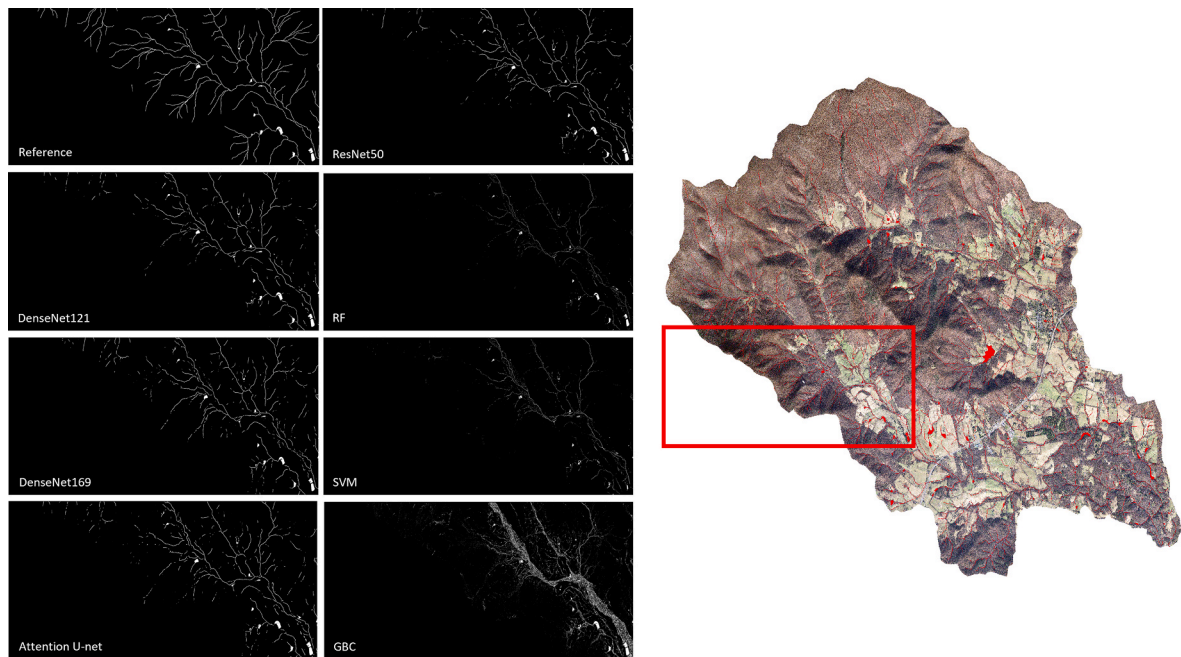


Fig. 8. Large-extent visualization of the prediction results for streamline delineation using transfer learning models (ResNet50, DenseNet121, DenseNet169, and Attention U-net) compared to traditional machine learning methods (Random Forest, SVM, and GBC). The transfer learning models produce smoother and more connected streamlines.

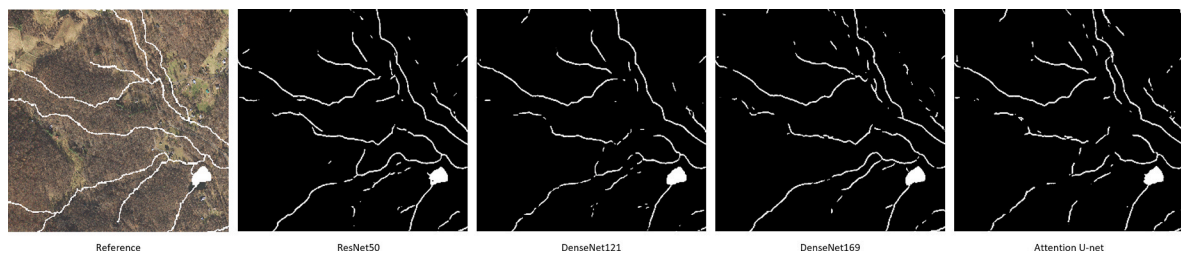


Fig. 9. Zoomed-in visualization of the prediction results for streamline delineation using transfer learning models (ResNet50, DenseNet121, DenseNet169, and Attention U-net). The ResNet50 model is particularly effective, producing cleaner and more well-connected channels compared to the other models.

4. Conclusions

This study advances hydrographic streamline delineation through two key contributions. Firstly, it has demonstrated the effectiveness of transfer learning using the U-net architecture with ImageNet pre-trained backbones, addressing challenges by leveraging extensive, generic ImageNet data for hydrographic tasks. The first experiment validated that transferring learning from image classification to streamline delineation outperformed models with randomly initialized weights, despite differences between tasks and datasets. It also highlights the enhanced transferability of the U-net model across geographically diverse areas. The second experiment showed that ImageNet-based transfer learning improved model adaptability from Rowan County to Covington, outperforming models trained from scratch. The research also evaluates various CNN architectures, finding (i) ResNet excels in capturing intricate details, (ii) DenseNet effectively captures fine and coarse features, (iii) Inception discerns features at multiple scales, and (iv) VGG networks capture low-level features but struggle with higher-order features. Furthermore, using ImageNet pre-trained backbones reduces training time and computational resources, highlighting transfer learning's efficiency and cost-effectiveness. Overall, fine-tuning U-net with ImageNet pre-trained weights proved more effective than training models from scratch, with recommendations to start with smaller models and evaluate larger ones as needed. Future research could

explore broader geographic transfers, additional CNN architectures, varied pre-training datasets, fine-tuning strategies, multi-task learning, meta-learning methods, and domain adaptation techniques to enhance model generalization and accuracy across diverse geographic areas.

Disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CRediT authorship contribution statement

Nattapon Jaroenchai: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Shaowen Wang:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Lawrence V. Stanislawski:** Writing – review & editing, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation. **Ethan Shavers:** Writing – review & editing, Validation, Investigation, Formal analysis. **Zhe Jiang:** Writing – review & editing, Validation, Investigation, Formal analysis. **Vasit Sagan:** Writing – review & editing, Investigation, Formal analysis. **E. Lynn**

Usery: Writing – review & editing, Validation, Investigation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset pivotal to the findings of this study is publicly available

via Figshare at <https://figshare.com/s/fe3744e17ae8378b1148>.

Acknowledgements

This study and associated materials are based in part upon work supported by the National Science Foundation (NSF) under grant numbers: 2112356, 2118329, and 2232860 as well as the Taylor Geospatial Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

Appendix A. Precision and Recall for Evaluation Metrics Corresponding to Table 3

	Initialized with ImageNet pre-trained weights			Initialized with random weights		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
DenseNet 169	87.91%	82.48%	85.11%	81.50%	82.10%	81.80%
ResNet 50	84.44%	83.11%	83.77%	77.60%	78.40%	78.00%
DenseNet 121	85.22%	82.28%	83.73%	82.00%	81.70%	81.85%
DenseNet 201	81.26%	84.73%	82.95%	80.10%	80.64%	80.37%
ResNet 34	85.54%	79.70%	82.51%	76.90%	76.46%	76.68%
ResNet 101	82.72%	80.86%	81.78%	75.60%	76.10%	75.85%
ResNet 152	80.07%	83.28%	81.65%	77.80%	79.00%	78.40%
Attention U-net	-	-	-	80.42%	82.17%	81.28%
VGG16	80.59%	81.37%	80.98%	73.70%	74.76%	74.23%
InceptionResNet V2	80.35%	81.18%	80.76%	73.50%	74.08%	73.79%
Inceptionv3	82.01%	78.89%	80.42%	75.30%	75.60%	75.45%
VGG19	79.98%	77.97%	78.96%	74.20%	75.38%	74.79%

Appendix B. Precision and Recall for Evaluation Metrics Corresponding to Table 4

	Covington			Rowan County		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ResNet50	67.33%	56.76%	61.59%	84.44%	83.11%	83.77%
DenseNet121	68.92%	46.70%	55.67%	85.22%	82.28%	83.73%
DenseNet169	74.36%	39.10%	51.25%	87.91%	82.48%	85.11%
Attention U-net	55.43%	43.02%	48.45%	80.42%	82.17%	81.28%

Appendix C. Precision and Recall for Evaluation Metrics Corresponding to Table 5

	Rowan County Fine-tuned to Covington (TL2)			Baseline in Covington		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
ResNet50	66.32%	64.86%	65.58%	67.33%	56.76%	61.59%
DenseNet121	75.66%	45.29%	56.66%	68.92%	46.70%	55.67%
DenseNet169	75.73%	42.90%	54.77%	74.36%	39.10%	51.25%
Attention U-net	63.70%	39.75%	48.95%	55.43%	43.02%	48.45%

Appendix D. Precision and Recall for Evaluation Metrics Corresponding to Table 6

Models/methods	Precision	Recall	F1-score in Covington
ResNet50	63.22%	68.02%	65.58%
DenseNet121	54.35%	59.08%	56.66%
DenseNet169	53.45%	56.14%	54.77%
Attention U-net	47.82%	50.12%	48.95%
Random Forest	44.27%	48.06%	46.14%
Support Vector Machine	38.33%	41.34%	39.81%
Gradient Boosting Classifier	28.15%	32.19%	30.09%

References

- Aggarwal, P., Mishra, N.K., Fatimah, B., Singh, P., Gupta, A., Joshi, S.D., 2022. COVID-19 image classification using deep learning: advances, challenges and opportunities. *Comput. Biol. Med.* 144, 105350 <https://doi.org/10.1016/j.combiomed.2022.105350>. Elsevier BV.
- Barnes, R., Callaghan, K.L., Wickert, A.D., 2021. Computing water flow through complex landscapes – Part 3: fill–Spill–Merge: flow routing in depression hierarchies. *Earth Surf. Dyn.* 9 (1), 105–121. <https://doi.org/10.5194/esurf-9-105-2021>. Copernicus GmbH.
- Bisong, E., 2019. Google colabatory. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress, pp. 59–64. https://doi.org/10.1007/978-1-4842-4470-8_7 [Software].
- Cao, H., Pan, Z., Chang, Q., Zhou, A., Wang, X., Sun, Z., 2021. Stream network modelling using remote sensing data in an alpine cold catchment. *Water* 13 (11), 1585. <https://doi.org/10.3390/w13111585>. MDPI AG.
- Chaurasia, A., Culurciello, E., 2017. LinkNet: exploiting encoder representations for efficient semantic segmentation. In: 2017 IEEE Visual Communications and Image Processing (VCIP). 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE. <https://doi.org/10.1109/vcip.2017.8305148>.
- Chen, Y., Fan, R., Yang, X., Wang, J., Latif, A., 2018. Extraction of urban water bodies from high-resolution remote-sensing imagery using deep learning. *Water* 10 (5), 585. <https://doi.org/10.3390/w10050585>. MDPI AG.
- Chen, Y., Tang, L., Kan, Z., Bilal, M., Li, Q., 2020. A novel water body extraction neural network (WBE-NN) for optical high-resolution multispectral imagery. *J. Hydrol.* 588, 125092 <https://doi.org/10.1016/j.jhydrol.2020.125092>. Elsevier BV.
- Chollet, F., others, 2015. Keras [Software] GitHub. Retrieved from. <https://github.com/fchollet/keras>.
- Colaco, S., Yoon, Y.J., Han, D.S., 2022. UIRNet: facial landmarks detection model with symmetric encoder-decoder. In: 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). IEEE. <https://doi.org/10.1109/icaaiic54071.2022.9722657>.
- Comer, P., Faber-Langendoen, D., Evans, R., Gawler, S., Josse, C., Kittel, G., Menard, S., Pyne, M., Reid, M., Schulz, K., Snow, K., Teague, J., 2003. *Ecological Systems of the United States, A Working Classification of U.S. Terrestrial Systems*. Arlington, Va., NatureServe, p. 75.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, Kai, Fei-Fei, Li, 2009. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). IEEE. <https://doi.org/10.1109/cvpr.2009.5206848>.
- Doneus, M., 2013. Openness as visualization technique for interpretative mapping of airborne lidar derived digital terrain models. *Rem. Sens.* 5 (12), 6427–6442. <https://doi.org/10.3390/rs5126427>. MDPI AG.
- Ehlschlaeger, C., 1989. Using the AT search algorithm to develop hydrologic models from digital elevation data. In: *Proceedings of International Geographic Information Systems (IGIS) Symposium*, Baltimore, MD, USA, pp. 275–281, 1989.
- Gao, J., Fan, W., Jiang, J., Han, J., 2008. Knowledge transfer via multiple model local structure mapping. In: *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08*. The 14th ACM SIGKDD International Conference. ACM Press. <https://doi.org/10.1145/1401890.1401928>.
- Gowroju, S., Kumar, S., Aarti, Ghimire, A., 2022. Deep neural network for accurate age group prediction through pupil using the optimized UNet model. In: Saini, D.K. (Ed.), *Mathematical Problems in Engineering*, 2022. Hindawi Limited, pp. 1–24. <https://doi.org/10.1155/2022/7813701>.
- Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., et al., 2019. Mapping the world's free-flowing rivers. *Nature* 569 (7755), 215–221. <https://doi.org/10.1038/s41586-019-1111-9>.
- Guptill, S.C., 1983. The role of digital cartographic data in the geosciences. *Comput. Geosci.* 9 (1), 23–26. [https://doi.org/10.1016/0098-3004\(83\)90032-8](https://doi.org/10.1016/0098-3004(83)90032-8). Elsevier BV.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Deep residual learning for image recognition (version 1). *arXiv*. <https://doi.org/10.48550/ARXIV.1512.03385>.
- Heidemann, H.K., 2012. Lidar base specification. In: *Techniques and Methods*. US Geological Survey. <https://doi.org/10.3133/tm11b4>.
- Hopkinson, C., Hayashi, M., Peddle, D., 2009. Comparing alpine watershed attributes from LiDAR, photogrammetric, and contour-based digital elevation models. *Hydrol. Process.* 23 (3), 451–463. <https://doi.org/10.1002/hyp.7155>. Wiley.
- Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2016. Densely connected convolutional networks (version 5). *arXiv*. <https://doi.org/10.48550/ARXIV.1608.06993>.
- Jasiewicz, J., Stepinski, T.F., 2013. Geomorphons — a pattern recognition approach to classification and mapping of landforms. In: *Geomorphology*, 182. Elsevier BV, pp. 147–156. <https://doi.org/10.1016/j.geomorph.2012.11.005>.
- Jiang, Z., He, W., Kirby, M.S., Sainju, A.M., Wang, S., Stanislawski, L.V., Shavers, E.J., Usery, E.L., 2022. Weakly supervised spatial deep learning for Earth image segmentation based on imperfect polyline labels. In: *ACM Transactions on Intelligent Systems and Technology*, 13. Association for Computing Machinery (ACM), pp. 1–20. <https://doi.org/10.1145/3480970>, 2.
- Kampffmeyer, M., Salberg, A.-B., Jenssen, R., 2016. Semantic segmentation of small objects and modelling of uncertainty in urban remote sensing images using deep convolutional neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE. <https://doi.org/10.1109/cvprw.2016.90>.
- Kirillov, A., He, K., Girshick, R., Dollár, P., 2017. A unified architecture for instance and semantic segmentation. Available at: <http://presentations.cocodataset.org/COCO17-Stuff-FAIR.pdf>.
- Lang, M., McDonough, O., McCarty, G., Oesterling, R., Wilen, B., 2012. Enhanced detection of wetland-stream connectivity using LiDAR. *Wetlands* 32 (3), 461–473. <https://doi.org/10.1007/s13157-012-0279-7>. Springer Science and Business Media LLC.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444. <https://doi.org/10.1038/nature14539>. Springer Science and Business Media LLC.
- Li, J., Li, T., Zhang, L., Sivakumar, B., Fu, X., Huang, Y., Bai, R., 2020. A D8-compatible high-efficient channel head recognition method. *Environ. Model. Software* 125, 104624. <https://doi.org/10.1016/j.envsoft.2020.104624>. Elsevier BV.
- Li, M., McGrath, H., Stefanakis, E., 2022. Geovisualization of hydrological flow in hexagonal grid systems. *Geographies* 2, 227–244. <https://doi.org/10.3390/geographies2020016>.
- Li, J., Wang, C., Xu, L., Wu, F., Zhang, H., Zhang, B., 2021. Multitemporal water extraction of dongting lake and poyang lake based on an automatic water extraction and dynamic monitoring framework. *Rem. Sens.* 13 (5), 865. <https://doi.org/10.3390/rs13050865>. MDPI AG.
- Li, X., Cen, M., Xu, J., Zhang, H., Xu, X.S., 2022. Improving feature extraction from histopathological images through a fine-tuning ImageNet model. *J. Pathol. Inf.* 13, 100115. <https://doi.org/10.1016/j.jpi.2022.100115>. Elsevier BV.
- Lin, P., Pan, M., Wood, E.F., Yamazaki, D., Allen, G.H., 2021. A new vector-based global river network dataset accounting for variable drainage density. *Sci. Data* 8 (1). <https://doi.org/10.1038/s41597-021-00819-9>. Springer Science and Business Media LLC.
- Liu, W., Chen, X., Ran, J., Liu, L., Wang, Q., Xin, L., Li, G., 2020. LaeNet: a novel lightweight multitask CNN for automatically extracting lake area and shoreline from remote sensing images. *Rem. Sens.* 13 (1), 56. <https://doi.org/10.3390/rs13010056>. MDPI AG.
- Lyu, F., Xu, Z., Ma, X., Wang, S., Li, Z., Wang, S., 2021. A vector-based method for drainage network analysis based on LiDAR data. *Comput. Geosci.* 156, 104892. <https://doi.org/10.1016/j.cageo.2021.104892>. Elsevier BV.
- Maggiore, E., Tarabalka, Y., Charpiat, G., Alliez, P., 2017. High-resolution aerial image labeling with convolutional neural networks. In: *IEEE Transactions on Geoscience and Remote Sensing*, 55. Institute of Electrical and Electronics Engineers (IEEE), pp. 7092–7103. <https://doi.org/10.1109/tgrs.2017.2740362>, 12.
- Marks, D., Dozier, J., Frew, J., 1984. Automated basin delineation from digital elevation data. *Geo-processing* 2 (3), 299–311.
- Martin, A., others, 2015. TensorFlow: large-scale machine learning on heterogeneous systems [Software]. <https://www.tensorflow.org>.
- Metz, M., Mitasova, H., Harmon, R.S., 2011. Efficient extraction of drainage networks from massive, radar-based elevation models with least cost path search. *Hydrol. Earth Syst. Sci.* 15, 667–678. <https://doi.org/10.5194/hess-15-667-2011>.
- Mukhlif, A.A., Al-Khateeb, B., Mohammed, M.A., 2023. Incorporating a novel dual transfer learning approach for medical images. *Sensors* 23 (2), 570. <https://doi.org/10.3390/s23020570>.
- O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. *Comput. Vis. Graph Image Process* 28 (3), 323–344. [https://doi.org/10.1016/s0734-189x\(84\)80011-0](https://doi.org/10.1016/s0734-189x(84)80011-0). Elsevier BV.
- Orlandini, S., Tarolli, P., Moretti, G., Dalla Fontana, G., 2011. On the prediction of channel heads in a complex alpine terrain using gridded elevation data. *Water Resour. Res.* 47 (2) <https://doi.org/10.1029/2010wr009648>. American Geophysical Union (AGU).
- Passalacqua, P., Do Trung, T., Foufoula-Georgiou, E., Sapiro, G., Dietrich, W.E., 2010. A geometric framework for channel network extraction from lidar: nonlinear diffusion and geodesic paths. *J. Geophys. Res.* 115 (F1) <https://doi.org/10.1029/2009jf001254>. American Geophysical Union (AGU).
- Poppenga, S.K., Gesch, D.B., Worstell, B.B., 2013. Hydrography change detection: the usefulness of surface channels derived from LiDAR DEMs for updating mapped hydrography 1. *JAWRA Journal of the American Water Resources Association* 49 (2), 371–389. <https://doi.org/10.1111/jawr.12027>. Wiley.
- Rajib, A., Golden, H.E., Lane, C.R., Wu, Q., 2020. Surface depression and wetland water storage improves major river basin hydrologic predictions. *Water Resour. Res.* 56 (7) <https://doi.org/10.1029/2019wr026561>. American Geophysical Union (AGU).
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: convolutional networks for biomedical image segmentation. In: *Lecture Notes in Computer Science*. Springer International Publishing, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
- Sangireddy, H., Stark, C.P., Kladzyk, A., Passalacqua, P., 2016. GeoNet: an open source software for the automatic and objective extraction of channel heads, channel network, and channel morphology from high resolution topography data. *Environ. Model. Software* 83, 58–73. <https://doi.org/10.1016/j.envsoft.2016.04.026>. Elsevier BV.
- Schmidhuber, J., 2015. Deep learning in neural networks: an overview. *Neural Network*. 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>. Elsevier BV.
- Shavers, E., Stanislawski, L., 2018. Streams do work: measuring the work of low-order streams on the landscape using point clouds. *Int. Arch. Photogram. Rem. Sens. Spatial Inf. Sci.* 42, 573–578. <https://doi.org/10.5194/isprs-archives-XLII-4-573-2018>.
- Shen, C., Chen, X., Laloy, E., 2021. Editorial: broadening the use of machine learning in hydrology. In: *Frontiers in Water*, 3. Frontiers Media SA. <https://doi.org/10.3389/frwa.2021.681023>.
- Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V., 2021. U-net and its variants for medical image segmentation: a review of theory and applications. *IEEE Access* 9,

- 82031–82057. <https://doi.org/10.1109/access.2021.3086020>. Institute of Electrical and Electronics Engineers (IEEE).
- Simley, J.D., Carswell Jr., W.J., 2009. The national map—hydrography. U.S. Geological Survey Fact Sheet 2009-3054 4. Available at: <https://pubs.usgs.gov/fs/2009/3054/>.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition (version 6). arXiv. <https://doi.org/10.48550/ARXIV.1409.1556>.
- Stanislowski, L.V., Shavers, E.J., Wang, S., Jiang, Z., Usery, E.L., Moak, E., Duffy, A., Schott, J., 2021. Extensibility of U-net neural network model for hydrographic feature extraction and implications for hydrologic modelling. *Rem. Sens.* 13 (12), 2368. <https://doi.org/10.3390/rs13122368>. MDPI AG.
- Stanislowski, L.V., Survila, K., Wendel, J., Liu, Y., Bittenfield, B.P., 2017. An open source high-performance solution to extract surface water drainage networks from diverse terrain conditions. *Cartogr. Geogr. Inf. Sci.* 45 (4), 319–328. <https://doi.org/10.1080/15230406.2017.1337524>. Informa UK Limited.
- Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, inception-ResNet and the impact of residual connections on learning (version 2). arXiv. <https://doi.org/10.48550/ARXIV.1602.07261>.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the inception architecture for computer vision (version 3). arXiv. <https://doi.org/10.48550/ARXIV.1512.00567>.
- Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. In: *Water Resources Research*, 33. American Geophysical Union (AGU), pp. 309–319. <https://doi.org/10.1029/96wr03137>, 2.
- Tarboton, D.G., Ames, D.P., 2001. Advances in the mapping of flow networks from digital elevation data. In: *Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges*, pp. 1–10. [https://doi.org/10.1061/40569\(2001\)166](https://doi.org/10.1061/40569(2001)166).
- Ünal, Z., Aktaş, H., 2023. Classification of hazelnut kernels with deep learning. *Postharvest Biol. Technol.* 197, 112225 <https://doi.org/10.1016/j.postharvbio.2022.112225>. Elsevier BV.
- Usery, E.L., 2011. The digital transition in cartography: USGS data innovations, 1970s. In: *Lecture Notes in Geoinformation and Cartography*. Springer Berlin Heidelberg, pp. 115–128. https://doi.org/10.1007/978-3-642-19088-9_7.
- Verdin, K.L., 2017. *Hydrologic Derivatives for Modelling and Analysis—A New Global High-Resolution Database (No. 1053)*. US Geological Survey.
- Wang, G., Kikuchi, Y., Yi, J., Zou, Q., Zhou, R., Guo, X., 2022. Transfer learning for retinal vascular disease detection: a pilot study with diabetic retinopathy and retinopathy of prematurity (version 1). arXiv. <https://doi.org/10.48550/ARXIV.2201.01250>.
- Wang, Z., Dai, Z., Poczos, B., Carbonell, J., 2019. Characterizing and avoiding negative transfer. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. <https://doi.org/10.1109/cvpr.2019.01155>.
- Wang, Z., Liu, J., Li, J., Meng, Y., Pokhrel, Y., Zhang, H., 2021. Basin-scale high-resolution extraction of drainage networks using 10-m Sentinel-2 imagery. In: *Remote Sensing of Environment*, 255. Elsevier BV, 112281. <https://doi.org/10.1016/j.rse.2020.112281>.
- Wilson, J.P., Aggett, G., Yongxin, D., Lam, C.S., 2008. Water in the landscape: a review of contemporary flow routing algorithms. *Advances in Digital Terrain Analysis*. https://doi.org/10.1007/978-3-540-77800-4_12. Springer.
- Xiao, Z., Du, M., Liu, J., Sun, E., Zhang, J., Gong, X., Chen, Z., 2023. EA-UNet based segmentation method for OCT image of uterine cavity. *Photonics* 10 (1), 73. <https://doi.org/10.3390/photonics10010073>. MDPI AG.
- Xu, Z., Guan, K., Casler, N., Peng, B., Wang, S., 2018. A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. *ISPRS J. Photogrammetry Remote Sens.* 144, 423–434. <https://doi.org/10.1016/j.isprsjprs.2018.08.005>. Elsevier BV.
- Xu, Z., Wang, S., Stanislowski, L.V., Jiang, Z., Jaroenchai, N., Sainju, A.M., Shavers, E., Usery, E.L., Chen, L., Li, Z., Su, B., 2021. An attention U-Net model for detection of fine-scale hydrologic streamlines. *Environ. Model. Software* 140, 104992. Elsevier BV. <https://doi.org/10.1016/j.envsoft.2021.104992>.
- Yakubovskiy, P., 2019. Segmentation models [Software] GitHub repository. Available at: https://github.com/qubvel/segmentation_models.
- Yakubovskiy, P., 2022. Tutorial — segmentation models 0.1.2 documentation. Readthedocs.io. <https://segmentation-models.readthedocs.io/en/latest/tutorial.html#training-with-non-rgb-data>.
- Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? ArXiv. <https://doi.org/10.48550/ARXIV.1411.1792>.
- Zhang, H., Loáiciga, H.A., Feng, L., He, J., Du, Q., 2021. Setting the flow accumulation threshold based on environmental and morphologic features to extract river networks from digital elevation models. In: *ISPRS International Journal of Geo-Information*, 10. MDPI AG, p. 186. <https://doi.org/10.3390/ijgi10030186>, 3.
- Zhang, Z., 2018. Improved adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). IEEE. <https://doi.org/10.1109/iwqos.2018.8624183>.
- Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2016. Pyramid scene parsing network (version 2). arXiv. <https://doi.org/10.48550/ARXIV.1612.01105>.
- Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote sensing: a comprehensive review and list of resources. In: *IEEE Geoscience and Remote Sensing Magazine*, 5. Institute of Electrical and Electronics Engineers (IEEE), pp. 8–36. <https://doi.org/10.1109/mgrs.2017.2762307>, 4.