Received: 5 February 2021 | Revised: 8 December 2021 | Accepted: 9 December 2021

DOI: 10.1002/rra.3927

SHORT COMMUNICATION

WILEY

Towards an automatic characterization of riverscape development by deep learning

Knut Alfredsen[1] | Arild Dalsgård[2] | Saeid Shamsaliei[2] | Jo Halvard Halleraker[1,3] | Odd Erik Gundersen[2]

[1]Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

[3]Norwegian Environment Agency, Trondheim, Norway

Correspondence
Knut Alfredsen, Department of Hydraulic and Environmental Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway.
Email: knut.alfredsen@ntnu.no

Funding information
Norwegian Research Council, Grant/Award Number: 289725

Abstract

Riverscapes are under pressure from anthropogenic development, and this challenges the conservation of biodiversity, hydromorphology and land types. To assess changes and restoration potential, an understanding of alteration to rivers overtime is necessary. This can be challenging due to lack of data, shortcomings in methods and data formats that are not easily incorporated into the assessment process. Historical aerial imagery exists for rivers prior to modification, but the manual classification is time-consuming. Deep learning is increasingly used in image processing, and here we outline how a convolutional neural network can be used to automatically classify black and white aerial imagery from the database of the Norwegian mapping authority into habitat types. It is demonstrated how historical imagery can be used to develop maps that can be processed further in a GIS to evaluate natural versus anthropogenic changes over time.

KEYWORDS
aerial imagery, deep learning, habitat classification, river

1 | INTRODUCTION

Pressures on rivers and adjacent landscapes due to increasing anthropogenic development alters rivers and the river environment and influences riverine and riparian habitat along rivers (Grill et al., 2019; Grizzetti et al., 2017; Wohl, 2019). Multiple stressors are currently influencing rivers such as altering flow patterns and creating barriers through hydropower development, restricting river channels by dredging and the development of flood management structures, gravel mining, road and railroad development and urbanization. These developments lead to habitat loss, fragmentation of ecosystems, a reduction in biodiversity and have impacts on the ecosystem services provided by rivers, therefore mitigation and restoration are proposed (Wohl, 2019). The need for restoration is further emphasised by the United Nations (UN) declaring the coming decade as the decade of ecosystem restoration (UN, 2020).

The EU Water Framework Directive (WFD) has been implemented to ensure sustainable use of river basin ecosystems and address restoration and mitigation measures where necessary to ensure healthy riverine ecosystems. Asa valuable basis for ecosystem-based management of water bodies in Norway and assessment of the restoration potential, historic aerial photos prior to land-use alterations are considered a valuable baseline. This is crucial to the design of relevant mitigation measures and ensures sustainable management in line with the best available management practise as proposed in the Norwegian white paper on biodiversity (Anon, 2016).

Understanding of the development and the importance of dynamics within a river landscape over time rely on the availability of data on historical states of the river. Such data can be difficult to utilise, and methods for pre-processing historical data into a format that is suitable for geographical analysis are therefore needed. Gurnell, Downward, and Jones (1994) investigated spatial and temporal planform changes in

the River Dee using historical maps in a GIS. Garcia, Dunesme, and Piegay (2019) used historical topographic maps and a map classification toolbox to extract river features from four different countries and assessed changes to the rivers over time. Using historical aerial imagery which in many cases is available from before the satellite age, is an interesting source for both spatial and temporal changes in river and floodplain features. Gurnell (1997) used aerial photographs to investigate channel changes over a period of 115 years from the River Dee using a GIS to analyse pictures. Zanoni, Gurnell, Drake, and Surian (2008) used a combination of topographic maps and a series of aerial photographs to study the development of a braided river in Italy. They georeferenced and then digitized river features for further hydromorphological analysis. For shorter rivers or small spatial scale studies, a manual approach may be feasible. However, for larger rivers or regional analysis, an automated procedure will be required to manage the large amounts of data in a realistic time scale.

Archives of historical aerial photos are available in many countries and for large parts of Norwegian river basins. However, they are rarely used as the basis for the assessment of rivers, and it can be time-consuming if manual segmentation and GIS-analysis are needed (Åström, Ødegaard, Hanssen, & Åstrøm, 2017; Bergan & Solem, 2018). The potential for a multitude of assessments related to changes in river structure both on a temporal and spatial scale is possible if automated classification methods could be applied.

The recent advancement of artificial intelligence has been attributed to the advances of deep learning (LeCun, Bengio, & Hinton, 2015), the rapid increase in computational power and the availability of large datasets. Deep learning has been shown to be powerful for perception tasks, and especially for image analysis. Krizhevsky, Sutskever, and Hinton (2017) used deep convolutional neural networks for classifying images and reduced the top-5 error rate of the ImageNet challenge from 26.1 to 15.3% (IMAGENET, 2012). Deep convolutional neural networks have also been used for image segmentation (Ronneberger, Fischer, & Brox, 2015). In the past, several image regional segmentation methods have been proposed including clustering, the Watershed algorithm, and graph-based algorithms. These methods find consistent regions or region boundaries. However, our goal is not only to subdivide images into different segments but also to determine the class of each segment. From which it follows that each pixel in an image is assigned a class label. In other words, our task is semantic segmentation (Zhang et al., 2018). We have investigated the use of traditional methods for solving the task, these methods were not able to provide an intuitive segmentation for the examples. While several methods to address the problems we encounter exist (Couprie, Najman, & Lecun, 2013; Pinheiro & Collobert, 2014; Ronneberger et al., 2015), these approaches required a considerable amount of fine-tuning. In addition, given that the deep convolutional networks are state-of-the-art in semantic segmentation (Long, Shelhamer, & Darrell, 2015) and outperform traditional models, we decide to proceed with deep learning approaches.

In this paper, we demonstrate the applicability of deep learning to determine river and riparian classes from historical aerial imagery. We present the network structure, training process, output data and their potential for analysis of changes in riverscapes. The method is suitable for processing large datasets, and since it operates on black and white images it can handle historical pictures taken prior to many of the most significant land-use changes. This is important for understanding the long historical development of river systems and how they should be managed in the future.

2 | MATERIALS AND METHODS

2.1 | Data

The Norwegian mapping authority provided a database of aerial imagery of the mainland of Norway (www.norgeibilder.no) covering both recent and historic pictures. Most historical datasets are in black and white, and we converted newer colour pictures to black and white to utilise the same algorithm for all datasets. All images are georeferenced in the database and we used images projected into EUREF89-UTM33N for all analysis.

2.2 | Data sets and study sites

The datasets used are shown in Table 1. Each source image downloaded from the Norwegian mapping authority has a size of 6,000 × 8,000 pixels. To prepare the images for the deep learning algorithms, they were subdivided into smaller images with a size of 512 × 512 pixels. Small images with more than 10% of unknown classes or only one single class were filtered out. As deep neural networks belong to supervised learning methods, they need labelled examples to learn. Hence, a subset of all images was manually annotated in a GIS system to form the initial training and the test datasets. Regions of the images were annotated as six different classes (Table 2). As there are differences in the intensity value of the source images, we normalized the intensity value with regards to average intensity value of all images in the training dataset. The test set was also normalized based on the average intensity of the training dataset. The pixels of class "unknown" was changed to the class of the nearest neighbour pixel. Data augmentation methods such as rotation with 90°, 180° and 270° as well as vertical and horizontal flips were applied and resulted in 11 additional images for each training image. Enhancing a

TABLE 1 Dataset used in training and testing of the developed model

Dataset	Number of images (512 × 512)	Augmented	Rivers
Initial	1,694	20,328	Gaula 1963, Lærdal 1976
Expanded	6,307	75,684	Surna 1963
Test	927	N/A	Selected images from Gaula 1963, Gaula 1998, Nea 1962

dataset in this way is common as it helps machine learning models to generalize.

The test set was designed to test how well the model was able to generalize. We selected three rivers. Nea 1962 was selected to test if the model was able to segment an unseen river, Gaula 1998 was selected to test if the model was able to segment the same river at different points in time, and then we used some different parts of Gaula 1963 that were not part of the initial dataset. The test set had been manually annotated by the same non-domain experts that annotated the initial dataset.

2.3 | Neural network architecture

For the purpose of this work, the deep convolutional neural network we use was a U-net model (Ronneberger et al., 2015) which has encoder-decoder architecture illustrated in Figure 1. The encoder section extracts features and downsamples the images. Then the decoder upsamples the results of the encoder and generates a segmentation mask of the same size as the input image.

A pre-trained VGG16 model (Simonyan & Zisserman, 2015) was used as the encoder. VGG16 contains 5 convolutional blocks that extract features of the input images. Encoder blocks down-sample the input which means decreasing the resolution of input and increasing the depth of it. This helps the blocks that receive more processed data (blocks that are further away from the input image) to extract more abstract information out of the input image. The decoder has five convolutional blocks to up-sample the output of the encoder. It means they decrease the depth of input and increase the resolution of it. This is because we need to classify each pixel, accordingly, the output should have the same resolution as input. However, the output of each encoder block is concatenated to the input of the corresponding decoder. It helps to recover the fine details when constructing the segmentation map in the decoder (Drozdzal, Vorontsov, Chartrand, Kadoury, & Pal, 2016). Since the VGG16 is pre-trained on RGB images which have 3 channels, in order to benefit from the pre-trained encoder, our one channel black and white images are copied to each of the 3 channels.

TABLE 2 Classes used for semantic segmentation of landscapes

Class	Description
W	Water covered areas
G	Gravel bars and point bars in the river—Vegetation free.
V	Forest and other vegetated areas in the riparian corridor
F	Farmland and cultivated land in the river corridor
H	Anthropogenic structures like houses and roads
U	Unknown

2.4 | Training procedure

The training of a deep convolutional neural network segmentation model was done in five steps: (a) train model, CNNi, on a manually annotated dataset (initial), (b) use CNNi to segment a new dataset, (c) manually correct errors made by CNNi on the new dataset, (d) train CNNe on the combined dataset (expanded) and (e) evaluate CNNe on the test dataset (test).

Initially, large images (8000 × 6000 pixels) from the River Gaula in 1963 and the River Lærdal in 1976 were manually annotated. This was done by individuals supervised by domain experts. The annotation covered areas within 200 m of the river. The large images were then divided into 512 × 512 images. Small images containing only one class along with those images containing more than 10% "unknown" class were filtered out. At this stage, 1,694 small images remained, but as deep learning methods are excessively hungry for data, we augmented the data so that we ended up with 20,328 images. This dataset, called initial in Table 1, was used for training an initial model, CNNi, using 80% as training data and 20% for testing.

After training the CNNi model, it was used to segment the SurnaRiver from 1963. The SurnaRiver dataset was then manually

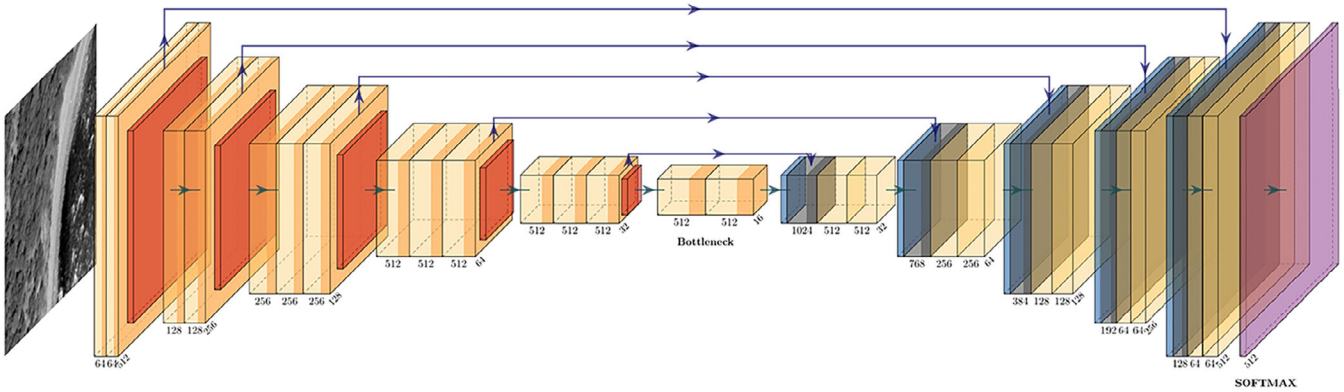

FIGURE 1 The model architecture. Light blue boxes are transpose convolution layers, dark blue boxes are concatenation between the skip connection from the encoder and the previous transpose convolution (right side of bottleneck), light orangeboxes are convolution layers and dark orangeboxes are max pooling layers (left side of bottleneck). The number under each box gives the number of channels for the output feature map. The tilted number at the end of each block is the image size for that block [Color figure can be viewed at wileyonlinelibrary.com]

corrected by non-domain experts so that this manually corrected dataset could be added to the initial dataset to form an expanded dataset containing 6,307 annotated images. This expanded dataset was augmented into 75,684 images and was used to train the second model, CNNe. In order to determine how many blocks of the encoder should be trained, we trained the model with different numbers of trainable blocks as a hyper-parameter using 80% of the expanded dataset as training data. We named each of the experiments using "freeze" as a prefix. Freeze 1 means the parameters of the first block of encoder will not change during the training and keep the initial value. Similarly, in freeze 2, the parameters of the first two blocks of encoder will remain constant at the training phase. Afterwards, validation data (the remaining 20% of the expanded dataset) was used to test and select the best number of trainable blocks as illustrated in Table 3. Mean Intersection over Union (MIoU, Equation 1) is used to measure the performance of each model.

$$\mathrm{MIoU} = \frac{1}{|C|}\sum_{c \in C}\left|\frac{\mathrm{pred}_c \cap \mathrm{true}_c}{\mathrm{pred}_c \cap \mathrm{true}_c}\right| \quad (1)$$

where C is the set of all the classes, pred_c is the predictions of the network on class c and true_c is the true label of class c.

The test datasets were evaluated using row and column normalized confusion matrices in addition to converting the predicted areas to polygons and doing a comparison between predicted types and the underlying aerial images using ArcMap GIS. The comparison of images was primarily done using visual inspection.

TABLE 3 Results of testing different configurations of the neural network on the validation set

Configuration	Valmiou	Valacc	Val loss
Freeze all	0.7006	0.8919	0.3102
Freeze first	0.7414	0.9064	0.2765
Freeze 3	0.7463	0.9078	0.2726
Freeze 2	0.7627	0.9107	0.2653
Freeze 1	0.7510	0.9076	0.2625
Freeze none	0.7513	0.9073	0.2661

Note: Different configurations are evaluated to select the best model.

3 | RESULTS

The model predictions on the test set for Gaula in 1963 are shown in the confusion matrix in Table 4, and a similar matrix for the test set from the River Nea is shown in Table 5. The model displays a consistent prediction of classes over all categories, except for gravel bars on the River Nea where the model predicts a large fraction of gravel as water or farmland. This is most likely due to difficult light conditions in the black and white photos making a distinction between the groups difficult.

The ability of the model to predict the river types is shown in Figure 2. There were minor errors in the prediction, for example, a small part of the inner part of the large gravel bar at the centre of the

TABLE 4 Row normalized confusion matrix for model predictions on test datasets for river Gaula in 1963

		Predicted class				
GAULA 1963		Water (%)	Gravel (%)	Vegetation (%)	Farmland (%)	Human (%)
True class	Water	91.31	0.38	1.38	6.93	0.00
	Gravel	7.84	76.73	6.72	6.10	2.60
	Vegetation	2.10	1.75	88.96	2.30	4.90
	Farmland	0.60	2.49	8.37	88.12	0.42
	Human	2.85	2.19	7.34	9.15	78.47

Note: Values on the diagonal represent the recall for that class.

TABLE 5 Row normalized confusion matrix for model predictions on test datasets for river Nea in 1962

		Predicted class				
NEA 1962		Water (%)	Gravel (%)	Vegetation (%)	Farmland (%)	Human (%)
True class	Water	95.36	0.14	1.83	2.39	0.28
	Gravel	22.68	53.15	8.04	10.07	6.05
	Vegetation	3.14	0.11	90.51	4.59	1.64
	Farmland	1.78	0.03	1.12	96.79	0.27
	Human	0.09	0.00	2.80	14.15	82.96

Note: Values on the diagonal represent the recall for that class.

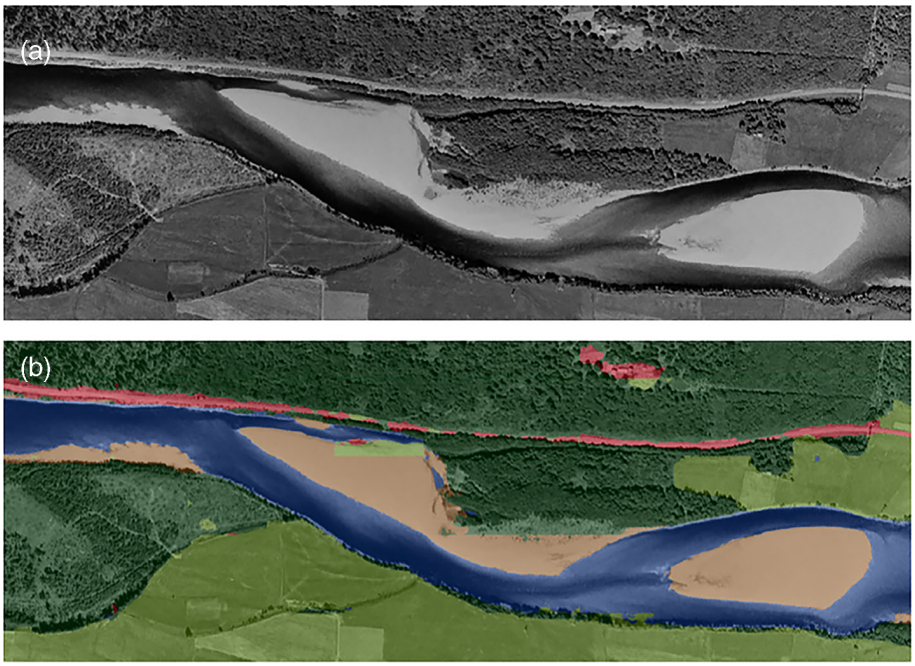

FIGURE 2 Predicted river types for a reach of Gaula in 1963. Panel a show the aerial imagery, panel b shows the image with overlaid classification. Blue—water, orange—man made, dark green—forest, light green—farmland and light brown—gravel. Flow direction is right to left [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 3 Predicted river types for a reach in Gaula in 1963. Panel a show the aerial image, panel b the image overlaid classification. Legend as for Figure 2. Flow direction is right to left [Color figure can be viewed at wileyonlinelibrary.com]

image was predicted as farmland, but in general, the predictions fit the image very well. Such errors were also easy to reclassify in the post-processing step. An example of a classification error is shown in Figure 3 where the model predicts farmland in the middle of the river.

As an example of an application of the method, we applied the neural network to the River Surnawhich was developed for hydropower in 1968. The images in Figure 4 show a section of the river that was bypassed by the water abstraction from the reservoir to the power plant. In Figure 4a the river was classified based on images from 1963, and Panel b) shows the same classification with overlaid aerial imagery from 2017. The loss of side channels through sedimentation and vegetation and a significant increase of vegetation on gravel bars which was clean before the regulation can be observed in Figure 4. This is a known effect of hydropower regulation and can be quantified through image analysis.

4 | DISCUSSION

In this paper, we have demonstrated how a deep convolutional neural network can be trained and used to assess the long-term alteration of riverscapes by automatic delineation of habitats from historical aerial black and white images. Black and white images pose an extra challenge since the segmentation is based on one channel of pixel intensity values. The proposed network was trained on aerial imagery manually annotated, and then applied to imagery from different years to create a database of habitats for each year. This database was the foundation for change analysis and can be combined with data on floods, ice runs and anthropogenic forcing's which forms the drivers behind changes in the structure of the river and the adjacent landscapes.

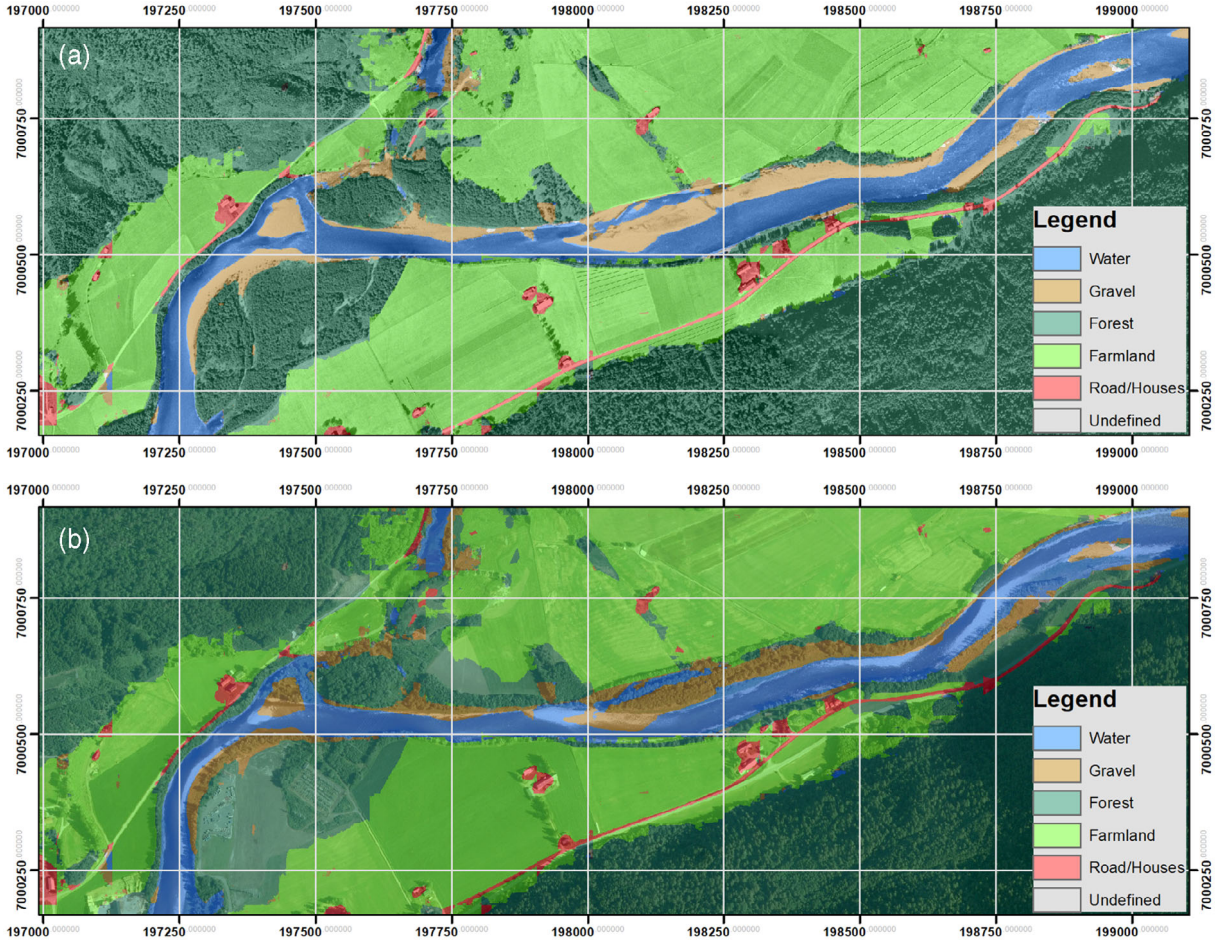

FIGURE 4 (a) Classification overlaid picture from Surna in 1963, (b) the 1963 classification overlaid an aerial picture from 2017 showing transformation of gravel bar to forest and narrowing of the river channel [Color figure can be viewed at wileyonlinelibrary.com]

We found the model to largely provide good predictions on the test datasets. The main exception was for gravel bars in the River Nea which we attribute to difficult contrasts in the images. An issue when the trained model was transferred to a new dataset was errors in classifications due to differences in light intensity between images in different datasets. In these cases, a light intensity adjustment was made to the images before training was carried out to improve the prediction. The model errors encountered were easy to correct during a GIS post-processing step. The combination of processing images with the neural network followed by post-processing was both simpler and faster than the time-consuming process of manual delineation of riverine characteristics. Since deep Learning approaches require a relatively large set of training data, datasets of annotated images were needed. Developing this dataset required significant manual effort, but we tried to reduce this effort by also utilizing enhanced data from model runs.

Reference conditions prior to major impacts are crucial for assessing ecological degradation, to classifying ecological status (Nybø & Evju, 2017), and designing ecologically well-functioning restoration measures (Guzelj, Hauer, & Egger, 2020) or identifying endangered riverscapes for ecosystem-based management (Aas, Indset, Prip, Platjouw, & Singsaas, 2020). Our methods enabled us to quantify river habitat development and to identify significant alterations by comparing pre- versus post-impact aerial photos. We found this method to be an efficient way to automatically classify large sets of images for evaluation of changes in river structure going back in time where only black and white images were available. We believe this approach also has a great potential for cost-efficient habitat assessment of larger areas and for refined classification for other habitat types, as archives of aerial photos are widespread as a basis for map-productions.

ACKNOWLEDGEMENT

The authors wish to thank Statkart—Geovekst for access to the aerial imagery and Oda, Mathea, Frøydisand Eleanor for the manual annotation of images. J. H. H. was funded by the Norwegian Research Council through the OFFPHD program, Grant No: 289725—Ecosystem-based management.

DATA AVAILABILITY STATEMENT

Aerial images of Norway are available from www.norgeibilder.no, copyright Statkart-Geovekst. Restrictions apply and the licenses may

be necessary to download images. Any other data can be requested from the authors.

ORCID

Knut Alfredsen https://orcid.org/0000-0002-4076-8351

REFERENCES

Aas, Ø., Indset, M., Prip, C., Platjouw, F.M. & Singsaas, F.T. (2020). Ecosystem-based manage-ment: Miracle or mirage? Mapping and rapid evidence assessment of international and Nordic research literature on ecosystem-based management Norwegian Institute for Nature Research. NINA report 1802.

Anon (2016). Natur for livet (Nature for life, text in Norwegian). Retrieved January 31, 2021, from https://www.regjeringen.no/no/dokumenter/meld.-st.-14-20152016/id2468099/.

Åström, J., Ødegaard, F., Hanssen, O. & Åstrøm, S. (2017). Endring i leveområder for elvesandjeger og stor elvebreddedderkopp ved Gaula. Forekomst og dynamikk av elveører fra 1947 til 2014. NINA Report 1314,32.

Bergan, M., & Solem, Ø. (2018). *Problemkartlegging, ungfiskovervåking og anslag på tapt areal i små sidevassdrag til Gaula*. Norwegian Institute of Nature Research 1497.

Couprie, C., Najman, L., & Lecun, Y. (2013). Learning hierarchical features for scene labeling. Pattern analysis and machine intelligence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*, 1915–1929. https://doi.org/10.1109/TPAMI.2012.231

Drozdzal, M., Vorontsov, E., Chartrand G., Kadoury, S. & Pal, C. (2016). The importance of skip connections in biomedical image segmentation. arXiv:1608.04117.

Garcia, J., Dunesme, S., & Piegay, H. (2019). Can we characterize river corridor evolution at a continental scale from historical topographic maps? A !Rst assessment from the comparison of four countries. *River Research and Applications*, *36*(6), 934–946. https://doi.org/10.1002/rra.3582

Grill, G., Lehner, B., Thieme, M., Geenen, B., Tickner, D., Antonelli, F., … Zarfl, C. (2019). Mapping the world's free-flowing rivers. *Nature*, *569*(7755), 215–221. https://doi.org/10.1038/s41586-019-1111-9

Grizzetti, B., Pistocchi, A., Liquete, C., Udias, A., Bouraroui, F., & van de Bund, W. (2017). Human pressures and ecological status of European rivers. *Scientific Reports*, *7*, 11. https://doi.org/10.1038/s41598-017-00324-3

Gurnell, A. (1997). Channel change on the river Dee meanders, 1946–1992, from the analysis of air photographs. *Regulated Rivers: Research and Management*, *13*(1), 13–26. https://doi.org/10.1002/(SICI)1099-1646(199701)13:1<13::AID-RRR420>3.0.CO;2-W

Gurnell, A., Downward, S., & Jones, R. (1994). Channel planform change on the river Dee meanders, 1876–1992. *Regulated Rivers: Research and Management*, *9*(4), 187–204. https://doi.org/10.1002/rrr.3450090402

Guzelj, M., Hauer, C., & Egger, G. (2020). The third dimension in river restoration: How anthropogenic disturbance changes boundary conditions for ecological mitigation. *Scientific Reports*, *10*, 13106.

IMAGENET. (2012). http://image-net.org/challenges/LSVRC/2012/.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

Long, J., Shelhamer, E. & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431 - 3440. Boston, MA. doi: 10.1109/CVPR.2015.7298965

Nybø, S., & Evju, M., Eds. (2017). An index-based assessment system for ecosystem condition in Norway-recommendation from an national expert commitee. (In Norwegian, English summary). Trondheim. https://www.regjeringen.no/no/dokument/rapportar-og-planar/id438817/.

Pinheiro, P. O. & Collobert, R. (2014). Recurrent Convolutional Neural Networks for Scene Labeling. ICML'14: Proceedings of the 31st International Conference on Machine Learning - 32, I-82–I-90.

Ronneberger, O., Fischer, P. & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. Paper presented at Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Springer, LNCS, 9351, 234–24.

Simonyan, K. and Zisserman, A. (2015). *Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations*, San Diego: ICLR.

UN. (2020). *UN Decade on Ecosystem Restoration*. Retrieved September 21, 2020, from https://www.decadeonrestoration.org.

Wohl, E. (2019). Forgotten legacies: Understanding and mitigating historical human alteration of river corridors. *Water Resources Research*, *55*, 5181–5201. https://doi.org/10.1029/2018WR024433

Zanoni, L., Gurnell, A., Drake, N., & Surian, N. (2008). Island dynamics in a braided river from analysis of historical maps and air photographs. *River Research and Applications*, *24*, 1141–1159. https://doi.org/10.1002/rra.1086

Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A. & Agrawal, A. (2018). Context encoding for semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 7151–7160. doi: 10.1109/CVPR.2018.00747.

How to cite this article: Alfredsen, K., Dalsgård, A., Shamsaliei, S., Halleraker, J. H., & Gundersen, O. E. (2021). Towards an automatic characterization of riverscape development by deep learning. *River Research and Applications*, 1–7. https://doi.org/10.1002/rra.3927