ORIGINAL ARTICLE

BathyNet: A Deep Neural Network for Water Depth Mapping from Multispectral Aerial Images

Gottfried Mandlburger[1] · Michael Kölle[2] · Hannes Nübel[2] · Uwe Soergel[2]

Abstract

Besides airborne laser bathymetry and multimedia photogrammetry, spectrally derived bathymetry provides a third optical method for deriving water depths. In this paper, we introduce BathyNet, an U-net like convolutional neural network, based on high-resolution, multispectral RGBC (red, green, blue, coastal blue) aerial images. The approach combines photogrammetric and radiometric methods: Preprocessing of the raw aerial images relies on strict ray tracing of the potentially oblique image rays, considering the intrinsic and extrinsic camera parameters. The actual depth estimation exploits the radiometric image content in a deep learning framework. 3D water surface and water bottom models derived from simultaneously captured laser bathymetry point clouds serve as reference and training data for both image preprocessing and actual depth estimation. As such, the approach highlights the benefits of jointly processing data from hybrid active and passive imaging sensors. The RGBC images and laser data of four groundwater supplied lakes around Augsburg, Germany, captured in April 2018 served as the basis for testing and validating the approach. With systematic depth biases less than 15 cm and a standard deviation of around 40 cm, the results satisfy the vertical accuracy limit Bc7 defined by the International Hydrographic Organization. Further improvements are anticipated by extending BathyNet to include a simultaneous semantic segmentation branch.

Keywords Spectrally derived bathymetry · Deep learning · Convolutional neural networks · Multispectral images · Airborne laser bathymetry

Zusammenfassung

BathyNet: Ein Deep Neural Network für die Wassertiefenkartierung aus multispektralen Luftbildern. Neben Flugzeuglaserscanning und Mehrmedienphotogrammetrie stellt die Analyse von Multispektralbildern eine dritte optische Methode zur Ableitung von Wassertiefen dar. In diesem Beitrag präsentieren wir BathyNet, ein an U-Net angelehntes Convolutional Neural Network, das auf hochauflösenden RGBC-Luftbildern (rot, grün, blau, coastal blue) basiert. Der Ansatz kombiniert photogrammetrische und radiometrische Methoden: Die Vorverarbeitung der originalen Luftbilder beruht auf Strahlenverfolgung der potentiell schrägen Bildstrahlen unter Berücksichtigung der inneren und äußeren Kameraorientierung. Die eigentliche Wassertiefenschätzung erfolgt danach durch Auswertung des radiometrischen Bildinhalts in einem Deep Learning Framework. 3D-Wasseroberflächen- und Wasserbodenmodelle, die aus simultan erfassten Laserbathymetrie-Punktwolken abgeleitet wurden, dienen als Referenz- und Trainingsdaten sowohl für die Bildvorverarbeitung als auch für die tatsächliche Tiefenschätzung. Damit unterstreicht der Ansatz die Vorteile einer gemeinsamen Verarbeitung der Daten von hybriden Sensorsystem bestehend aus Laserscannern und Kameras. Für die Überprüfung und Validierung des Ansatzes wurden RGBC-Bilder und Laserscandaten von vier grundwassergespeisten Seen rund um Augsburg, Deutschland, herangezogen. Mit einem systematischen Tiefenfehler von weniger als 15 cm und einer Standardabweichung von etwa 40 cm entsprechen die erzielten Ergebnisse der Genauigkeitsklasse Bc7 des von der International Hydrographic Organization definierten Standards für hydrographische Vermessungen. Weitere Verbesserungen sind durch die Erweiterung von BathyNet um simultane sematische Segmentierung zu erwarten.

✉ Gottfried Mandlburger
gottfried.mandlburger@geo.tuwien.ac.at

[1] Department of Geodesy and Geoinformation, TU Wien, Wiedner Hauptstr. 10-12, 1040 Vienna, Austria

[2] Institute for Photogrammetry, University of Stuttgart, Geschwister-Scholl-Str. 24D, 70174 Stuttgart, Germany

1 Introduction

Regular monitoring of water depths (i.e., bathymetry) is gaining ever more importance in the context of climate change and human interference in the sensitive littoral zone for both coastal areas and inland waters. On the coastal side, monitoring is required for documenting the impact of natural hazard events like earthquake sea waves or hurricanes (Salameh et al. 2019; Bergsma et al. 2019), changes in benthic habitats (Parrish et al. 2016; Eugenio et al. 2017; Purkis et al. 2019), and sea level rise (Sam et al. 2018; Rasquin et al. 2020). For inland waters, remote sensing based bathymetric surveys are used to measure the depth of relatively clear and shallow rivers and lakes (Hilldale and Raff 2008; Kasvi et al. 2019) as the basis for monitoring flood extents, fluvial morphodynamics, and in-stream habitats (Mandlburger et al. 2015; Carrivick and Smith 2019). The importance of tracing fresh water resources on a trans-national level is also emphasized by three European Directives, namely the Water Framework Directive (European Union 2000), the Flood Directive (European Union 2007), and the Fauna-Flora-Habitat Directive (European Union 1992).

While deeper areas beyond a water depth of approx. 20 m require sonar techniques for capturing underwater topography, optical methods are well suited for data acquisition of the shallow littoral zone. In particular, airborne data acquisition based on images or laser scans provide a seamless transition from under water via the water-land boundary to the dry part of the surrounding alluvial or coastal area. The three main optical remote sensing methods for deriving bathymetry (Kasvi et al. 2019) are: airborne laser bathymetry (ALB), also referred to as airborne laser hydrography (Philpot 2019; Guenther et al. 2000), multimedia photogrammetry, also referred to as photobathymetry (Westaway et al. 2001; Maas 2015; Dietrich 2016; Mandlburger 2019), and spectrally derived bathymetry (Lyzenga 1978; Lyzenga et al. 2006; Legleiter et al. 2009; Misra and Ramakrishnan 2020; Hodúl et al. 2018). While the achievable depth of multimedia photogrammetry is inherently restricted to the Secchi depth (Effler 1988) and can reach about 1.5 times the Secchi depth for spectral methods (Gao 2009), the maximum depth performance of ALB sensors is around three times the Secchi depth (Mandlburger 2020).

Today, industrial cameras are available at a much lower price compared to bathymetric laser scanners. Such cameras can be mounted on manned or unmanned airborne platforms (Agrafiotis et al. 2019). For satellite remote sensing, the ICESat-2 mission provides bathymetric LiDAR (Light Detection And Ranging) data with global coverage every 91 days (Parrish et al. 2019; Ma et al. 2020), but the spatial resolution (across track: 280 m) is lower compared to very high-resolution satellite image sensors (e.g., WorldView-3,

Landsat 8, Pleiades, QuickBird, RapidEye, etc.) used for multispectral approaches (Sagawa et al. 2019; Cahalane et al. 2019). In summary, although laser bathymetry provides better depth performance in general, image based techniques are frequently employed due to lower costs and better availability of image data.

The recent generation of airborne bathymetric sensors include laser scanners and multispectral cameras on the same platform. This configuration enables joint processing of data from both active and passive sensors. However, to date there are few studies exploiting the benefits of concurrent scan and image data acquisition. Considering the rapid rise of machine learning in general and convolutional neural networks in particular, hybrid sensors open up the possibility of calibrating bathymetric models and products derived from multispectral images with concurrently acquired laser data as reference. This strategy is especially beneficial for deep learning approaches, which often require abundant training data.

In this article, we therefore propose a comprehensive framework for deriving bathymetry from multispectral aerial images using a deep neural network (DNN) based on concurrently acquired laser bathymetry data used for training, testing, and validation. Our approach combines strict geometric and photogrammetric preprocessing of the wide-angle aerial images and actual bathymetry estimation based on the radiometric content of RGB images and an additional coastal blue wavelength (RGBC) using an U-Net Convolution Neural Network (CNN) (Ronneberger et al. 2015). In a first step, the laser bathymetry point clouds are classified into water surface, dry and wet ground, and other (vegetation, buildings, water column, etc.). The water surface points are used to generate a Digital Water Surface Model (DWSM) which, in turn, serves as the basis for laser pulse travel time and refraction correction (Westfeld et al. 2017) of the water bottom points. The corrected bottom points comprise the input for interpolating a Digital Terrain Model (DTM) of the underwater topography. Next, the precise (interior and exterior) image orientations are derived within a standard Structure-from-Motion workflow. We then intersect each image ray with the DWSM and the refracted image ray with the DTM. This results in the slanted in-water distance that serves as the basis for running the U-Net CNN with the multispectral RGBC information as input channels. The network delivers water depth estimates for every pixel. The CNN is then applied to unseen data for testing, and the resulting point clouds are subject to further post-processing where the high overlap of 90% is exploited for median-based filtering of the final product.

Our main contribution is, therefore, the combination of rigorous geometric and deep learning based radiometric data processing. In the absence of concurrently acquired laser data, the trained net can be used for pure radiometry based

depth estimation. In all other cases, the simultaneously captured laser data serve for fine tuning the pre-trained models when applied to water bodies with different characteristics. To a certain extent, the approach is also capable of bridging data voids in areas beyond the maximum penetration of the LiDAR and, therefore, increases the completeness of the derived underwater DTM. This expectation is fuelled by literature reports stating a depth penetration of 1.5–2 times the Secchi depth using spectral methods (Gao 2009; Duplančić Leder et al. 2019) whereas high-resolution topo-bathymetric laser scanners only achieve 1–1.5 Secchi depths (Mandlburger 2020).

While traditional, regression based methods for spectrally derived bathymetry are simpler to use and the derived explicit model parameters are interpretable, we opt for a CNN-based approach, as CNNs (i) can approximate arbitrary functions and (ii) incorporate spatial context. This suggests that CNN based depth retrieval might outperform traditional optical inversion methods by using all input channels of multispectral images in parallel for analysing the complex backscatter of solar radiation interacting with the water surface, column, and bottom.

The remainder of the article is structured as follows: Sect. 2 provides a brief review of SDB related publications with a specific focus on machine learning techniques. In Sect. 3, the study area around Augsburg, Germany is introduced together with the image and laser datasets captured in April 2018. The area comprises a dozen groundwater supplied artificial lakes located within the Lech River floodplain. Section 4 contains a detailed description of our BathyNet depth retrieval framework. In Sect. 5, we show the results of four selected lakes including network performance and error metrics. Section 6 critically discusses the achieved results and also highlights the limitations of the approach. Section 7, finally, summarizes the main findings and conclusions together with an outlook on future work.

2 Related Work

The derivation of water depths based on multispectral images has a long tradition with the fundamental physical models described in Lyzenga (1978), Stumpf et al. (2003), and Lyzenga et al. (2006). The most prominent parameter concerning light interaction with water is the wavelength dependent exponential decay of the radiance within the water column due to scattering and absorption, which is why most of the models use logarithmic expressions to establish the relation between reflectance and depth. To minimize the influence of varying bottom reflectivity, the individual radiometric channels (coastal blue, blue, green, red) of multispectral images are used within the depth estimation procedure. In this context, the Optimal Band Ratio

Analysis (OBRA) algorithm has gained broad attention in the SDB literature (Legleiter et al. 2009).

In the recent years, several studies have been published based upon the above mentioned publications: Salameh et al. (2019) provide a review of spaceborne remote sensing for monitoring beach topography and nearshore bathymetry. Muzirafuti et al. (2020) use log-band ratio and OBRA methods for deriving shallow water bathymetry from multispectral QuickBird satellite images. Rossi et al. (2020) apply the methods of Lyzenga et al. (2006) and Stumpf et al. (2003) to multispectral Unmanned Aerial Vehicle (UAV) images featuring the same spectral bands as the WorldView-2 satellite sensor. For mapping bathymetry of inland running waters, Gentile et al. (2016) use UAV hyperspectral images and apply empirical models for depth retrieval, which are applicable under a range of specific field conditions. Starek and Giessel (2017) combine multimedia photogrammetry and band ratio based optical inversion for generating seamless topo-bathymetric elevation models. Bué et al. (2020) focus on the intertidal zone and even include the IR band, which exhibits a high attenuation coefficient in water, for mapping very shallow water bathymetry. They propose a logistic regression approach and estimate three parameters of a sigmoid function based on the IR channels of Sentinel-2 images.

Hernandez and Armstrong (2016) employ WorldView-2 multispectral images to derive bathymetry of a coastal scene and validate their results against bathymetric maps from ship acoustic surveys and ALB. While the authors invested in thorough preprocessing of the image data based on Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) and the Cloud Shadow Approach (CSA) for atmospheric corrections, they only apply a simple band ratio model as basis for their colour-to-depth regression. Zhang et al. (2020) developed an Inherent Optical Parameters Linear Model (IOPLM) to estimate shallow water bathymetry of a coastal scene based on WorldView-2 and Sentinel-2A multispectral images. Cahalane et al. (2019) compare different satellite datasets (Landsat 8, RapidEye, and Pleiades) and use a log ratio transform to derive bathymetry. Legleiter and Fosness (2019), finally, discuss the limits of spectrally based depth mapping for rivers containing both optically deep and shallow areas. They developed a technique for inferring the maximum detectable depth based on their OBRA algorithm utilizing multibeam sonar data and hyperspectral images of a large and deep river. They furthermore discuss the portability of the depth-reflectance relation calibrated at a certain section of the river to other sites along the same river and underline the importance of model calibration incorporating a broad range of water depths.

All cited articles so far employ SDB approaches based on traditional data processing pipelines. In the following, we

briefly review contributions using machine learning (ML). Agrafiotis et al. (2019), Agrafiotis et al. (2020) apply a Support Vector Regression (SVR) framework to compensate for the refraction effect on 3D underwater point clouds obtained from raw Structure-from-Motion (SfM) and Multi-View Stereo (MVS) processing pipelines. The authors validated their approach with dense bathymetric LiDAR point clouds and terrestrial surveys with a total station as ground truth and achieved mean deviations of 0–20 cm as well as a standard deviation of 10–40 cm, which is close to the Special Order accuracy limit defined by the International Hydrographic Organization (IHO) of ± 25 cm (IHO 2020). While their approach does not exploit the multispectral image content for actual depth estimation in the first place, it constitutes a combination of geometric-photogrammetric data processing and machine learning, an idea, which is also present in our approach. Sagawa et al. (2019), in contrast, used SVR directly on multispectral Landsat 8 images to solve a convex optimisation problem for estimating multi-temporal bathymetry of a coastal scene. Compared to echo sounding as reference, they achieved a Mean Absolute Error (MAE) in the range of 1 m.

Sagawa et al. (2019) apply Random Forest machine learning and multi-temporal Landsat 8 satellite images to create a generalized depth estimation model. Their approach reached a Root Mean Square Error (RMSE) of 1.4 m for water depths between 0 and 20 m compared to bathymetric LiDAR and Single Beam Echo Sounding (SBES) as reference. Misra et al. (2018) use a nonlinear Support Vector Machine (SVM) machine learning technique based on Multi Beam Echo Sounding (MBES) data as reference and Landsat 7 and 8 multispectral images for bathymetry estimation along Sint Maarten Island and Ameland Inlet, The Netherlands. Splitting the MBES reference data in 80% for training and 20% for validation, they achieve a depth error of 8–14.5% of the water depth (maximum penetration depth: 15 m). The authors report that the SVM results outperform the conventional linear as well as the blue-green band ratio transform models. As the final example in this group, Legleiter and Harrison (2019) applied a generalized version of the ORBA algorithm and a K-nearest neighbour machine learning algorithm to multispectral satellite and hyperspectral UAV images. The authors found that the K-nearest neighbours algorithm showed higher observed-versus-predicted R^2 values and that preprocessing of satellite images was not necessary for depth retrieval. The latter is considered an important general statement for machine learning based depth retrieval, as the models tend to learn the required correction parameters themselves. This especially applies to deep networks like, e.g. Convolutional Neural Networks (CNN), which are able to approximate arbitrary functions (Zhou 2020). As our contribution is in its core a CNN

approach, we now review recent papers employing neural networks for bathymetry estimation.

Makboul et al. (2017) use the logarithms of Landsat 8 band reflectance values as input for a simple Artificial Neural Network (ANN) with one hidden layer to estimate the bathymetry of a harbor area. After correcting the images for atmospheric and sun glint effects, the first three bands (coastal blue, blue, green) were employed to train the ANN relying on Global Positioning System (GPS) and SBES measurements as depth reference. A similarity to our approach with respect to the data basis is the use of the water penetrating coastal blue wavelength. The authors report increased R^2 values of the ANN compared to conventional log-ratio based depth estimation.

Li et al. (2018) propose a deep convolution neural network named (DeepUNet) for sea-land separation of high-resolution optical remote sensing images. They compared their classification result with other convolutional networks (original U-Net, SegNet, SeNet) and report an accuracy gain of 1–2%. As in our approach, they apply a U-Net framework, but DeepUNet is not aiming at estimating bathymetry. Dickens and Armstrong (2019) combine DeepUNet for coastline detection with a Recurrent Neural Network (RNN) to predict water depth using Orbview-3 images, preprocessed depth reference data for training, and nautical charts for validation. They achieve a MAE in the meter range, which fails to meet the IHO standards 1a, 1b, and Special Order (IHO 2020). Liu et al. (2018) propose a locally adaptive back-propagation neural network (LABPNN) for depth retrieval. Their approach extends regular back-propagation networks and deals with non-stationarity by estimating water depth with an ensemble of LABPNNs, which are inversely weighted by the distance of an estimation point to the next training sample. The authors tested their method with multispectral data from different satellite sensors (IKONOS-2, Landsat 7) and report an accuracy increase of 5–7% compared to regression based depth retrieval.

Ma et al. (2020) employ bathymetric LiDAR data from the ICESat-2 ATLAS (Advanced Topographic Laser Altimeter) instrument (Parrish et al. 2019) as reference for classical regression based depth estimation of Sentinel-2 multispectral images. The main advantage of such an approach is that both image and depth reference data acquisition takes place with only a short time delay due to the high revisit cycles, especially of Sentinel-2. In our work, we even go beyond this idea by employing simultaneously captured laser bathymetry data as the reference for our CNN-based spectrally derived bathymetry approach. The idea of using concurrently captured reference data is also pursued by Slocum et al. (2020). Their approach bases on reference points derived via multimedia photogrammetry (Maas 2015) for regression or shallow neural network based depth estimation. While the majority of SDB approaches rely on ortho-rectified images,

the approach of Slocum et al. (2020) uses the raw, perspective aerial images as basis and employ ray-tracing and proper refraction correction in their processing pipeline. We also believe that this is the best choice, especially for wide-angle images commonly used in manned and unmanned aerial photography. In that sense, their overall processing strategy is similar to our approach. As stated above and in contrast to Slocum et al. (2020), our approach incorporates concurrent laser bathymetry point clouds as reference. This exhibits the following main advantages: (i) the accuracy level of laser bathymetry is generally higher compared to multimedia photogrammetry as, e.g., reported by Maas (2015) and Mandlburger (2019), (ii) laser bathymetry depth performance is usually beyond the visual depth (Philpot 2019) while multimedia based techniques are inherently restricted to the Secchi depth, and (iii) laser bathymetry does not rely on image texture, which is especially important in areas featuring very homogeneous bottom type (e.g., sand). Apart from that, the approach of Slocum et al. (2020) differs from ours concerning the underlying SDB method. While Slocum et al. (2020) focus on the photogrammetric part of the processing pipeline, we combine geometric pre- and postprocessing with state-of-the-art CNN-based depth estimation.

3 Materials

3.1 Study Area

The study area is located in the floodplain of the Lech river east of the City of Augsburg, Bavaria, Germany (48° 22′ N, 10°53 ′ E). The Lech is a 255 km long alpine river, which deposited gravel sediment in the floodplain area around Augsburg. Open pit mining of gravel led to the formation of around a dozen groundwater supplied clear water lakes within the floodplain. Four of these artificial lakes (Autobahnsee, Helenensee, Friedberger See, and Kuhsee) served as specific study sites for developing the proposed BathyNet water depth retrieval approach and for validating the results. Figure 1 shows an overview map of the study area (Fig. 1a) and RGB orthophoto maps of the four lakes (Fig. 1b–e).
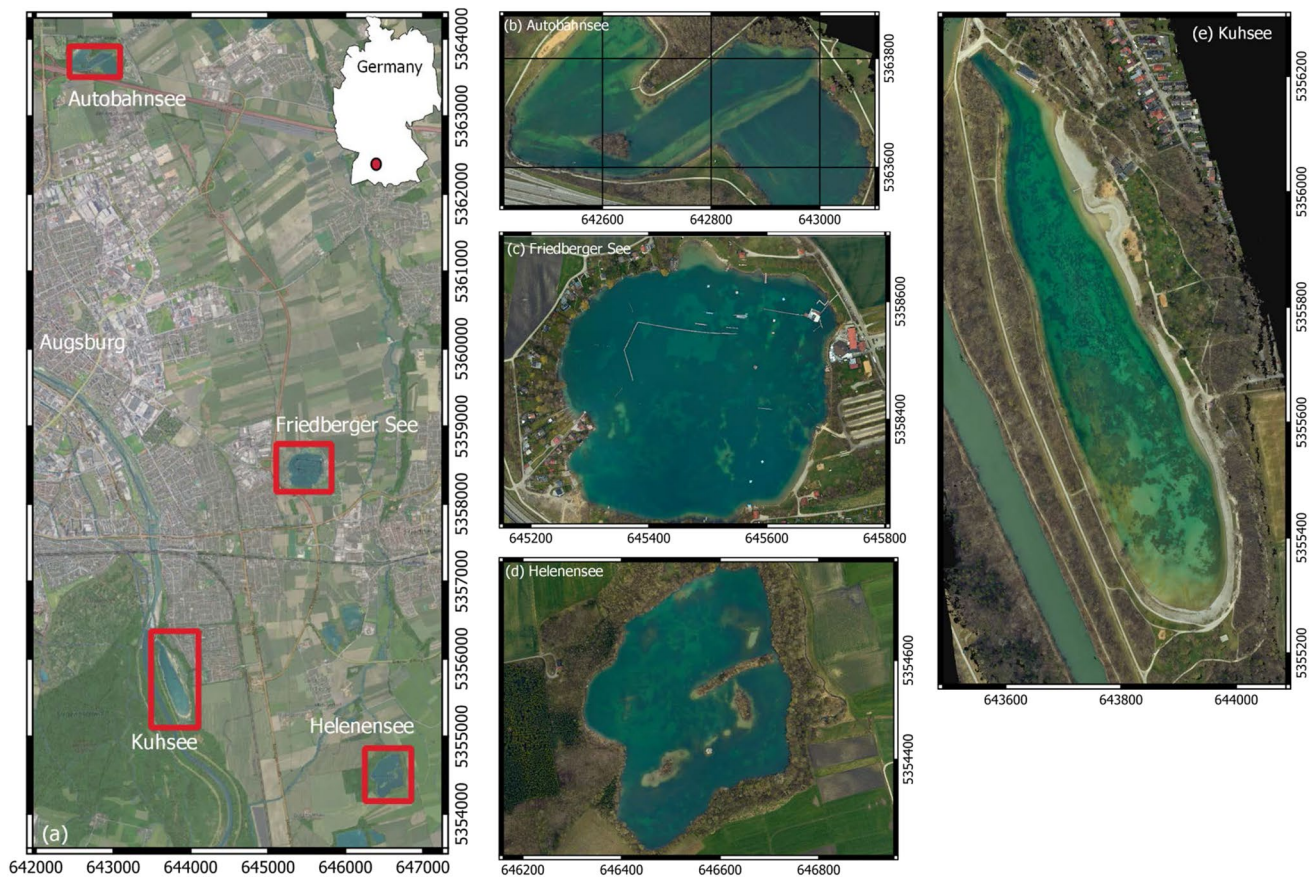
Fig. 1 Study area Augsburg, Bavaria, Germany; **a** overview map including location of study area within Germany and selected lakes marked with red rectangles; Orthophoto maps of Autobahnsee (**b**), Friedberger See (**c**), Helenensee (**d**), and Kuhsee (**e**); Coordinate Reference System: ETRS89/UTM32 (EPSG:25832)

3.2 Datasets

To test the feasibility of BathyNet for deriving high-resolution shallow water bathymetry, an airborne data acquisition was carried out on April, 9th, 2018. The hybrid sensor system comprised two IGI DigiCAM 100 cameras (pixel size: $4.6 \times 4.6 \mu m^2$, image size: 11608×8708 pixel) and a RIEGL VQ-880-G topo-bathymetric laser scanner (pulse repetition rate: 550 kHz, laser beam divergence: 1.1 mrad). The imaging system consisted of a standard RGB camera and a pan-chromatic sensor, which was restricted to Coastal Blue ($\lambda = 400$ to 460 nm) radiation with a filter mounted in front of the lens. Both cameras were equipped with a 50 mm wide-angle lens. A more detailed description of the camera system including the transmission curves of the individual spectral channels is provided in Mandlburger et al. (2018).

The captured data set comprised around a dozen of groundwater supplied lakes, form which the four lakes depicted in Fig. 1 were chosen as specific test sites due to their favourable water transparency. With a flying height of approximately 600 m above ground level (AGL), the ground sampling distance (GSD) of the RGB/Coastal Blue images was 5 cm and the image footprints on the ground were $600 \times 450 \, m^2$. Due to the limited size of the captured water bodies, the images generally featured enough dry land areas to enable proper image orientation. In total, approximately 20 flight strips were captured with an along-strip image overlap of 90%. The high redundancy is exploited in data postprocessing for smoothing the depth estimates and filtering of random measurement errors.

The LiDAR penetration depth varied between 1.5 and 7.5 m depending on water turbidity and bottom reflectance. Full water bottom point coverage was achieved for three of the four selected lakes (Autobahnsee, Helenensee, Kuhsee), and only Friedberger See contained an area with water depths

larger than 8 m beyond the penetration depth of the VQ-880-G sensor, which features a claimed maximum depth penetration of 1.5 times the Secchi depth. In addition to the airborne data, 12 black and white checker board targets were measured at Friedberger See and Autobahnsee with a Trimble RTK GNSS (Global Navigation Satellite System) receiver. The checker board targets served as control points for image bundle block adjustment.

While the captured laser bathymetry point clouds mainly act as reference data for training, testing, and validation, the multispectral images constitute the main source for our deep learning based depth retrieval approach. As a preliminary processing step, the simultaneously exposed coastal blue and RGB images were merged to a single 4-band-RGBC image using homography (Szeliski 2011). All remaining processing steps are described in Sect. 4.

4 Methods

This section describes the details of our approach starting with a coarse outline of the data processing pipeline and some underlying rationale (Sect. 4.1), followed by an in-depth description of the main processing stages, namely reference data generation (Sect. 4.2), actual CNN based depth retrieval (Sect. 4.3), and 3D point cloud derivation thereafter (Sect. 4.4).

4.1 Overview

Lyzenga et al. (2006) formulate a simple physically based SDB model, which is the starting point for many other contributions in the field. According to this model, the relation between subsurface reflectance $R(h)$ and water depth h can be written as:

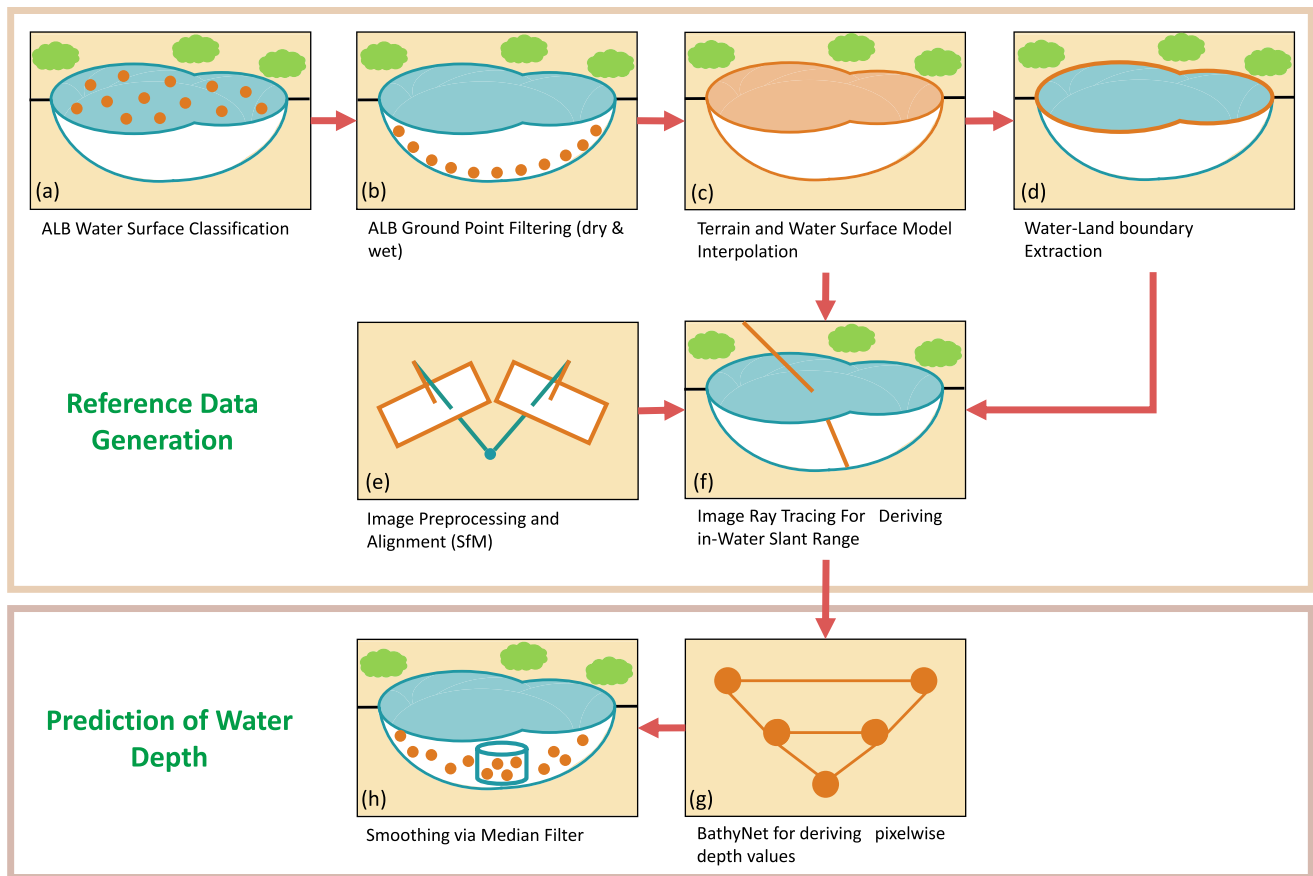**Fig. 2** Conceptual sketch of aerial image ray paths interacting with water
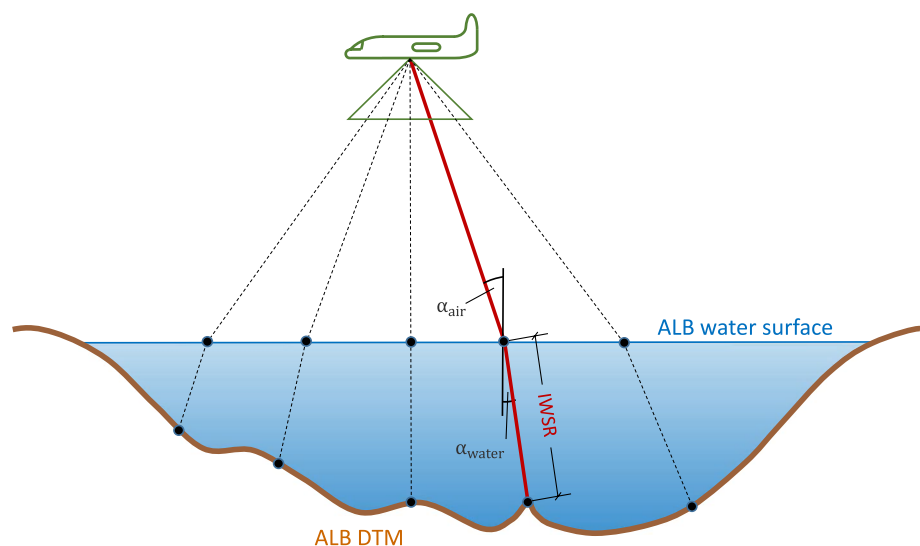
Fig. 3 Overall BathyNet depth retrieval workflow

$$R(h) = r_v + (r_b - r_v)e^{-\alpha h} \tag{1}$$

Parameter r_v denotes the reflectance due to volume scattering of an effectively infinite water column, r_b is the bottom reflectance, and α the effective attenuation coefficient (sum of coefficients for upwelling and downwelling light). Equation 1 represents an approximation to the radiative transfer solution and directly relates wavelength dependent radiance with depth h. This, however, is only applicable for nadir images with a sufficiently small opening angle, which is given for many high-resolution optical satellite images. In all other cases, the radiometric information stored in the image pixels relates to the slant distance between the intersection point of an image ray with the water surface and the corresponding second intersection of the refracted image ray in water with the water bottom. Slocum et al. (2020) coined the term in-water slant range (IWSR), which we adopted in this manuscript. Figure 2 illustrates the path of individual image rays refracted at the water surface and reflected at the water bottom. Calculation of the slant ranges for arbitrary image pixels requires continuous models of the water surface

and bottom, which in our approach are derived from concurrently captured laser bathymetry data.

Figure 3 illustrates the overall depth estimation workflow. The upper box is dedicated to data preprocessing and reference data generation, while the lower box sketches pixelwise depth retrieval and data postprecessing including 3D point cloud generation and median filtering of the predicted water bottom exploiting the high image overlap. Reference data generation comprises classification of the laser bathymetry point cloud into water surface, water bottom and (dry) ground points, the interpolation of gridded digital water surface and terrain models (DWSM, DTM), and water-land boundary (WLB) derivation. After merging the separate coastal blue channel with the RGB bands and image alignment based on SfM, the image rays are intersected with the DWSM, and the bent, sub-aqueous rays are intersected with the DTM providing the reference IWSR as the primary output. The RGBC channels together with the corresponding IWSR constitute the input for the BathyNet CNN. After training and testing the network, the model is applied to unseen data. BathyNet basically predicts IWSR

values (in image space), which are transformed into a 3D point cloud by reconstructing the 3D image rays based on the exterior and interior image orientation parameters and by performing ray tracing with the water surface, followed by refraction correction. The 3D bottom points are finally calculated by multiplying the refracted image unit ray vector with the predicted IWSR. Smoothing of the final water bottom model can be achieved by median filtering the 3D point clouds of overlapping images. While the standard deviation of the bottom point heights derived from individual images constitutes a measure of precision, the mean error (ME) calculated as the difference between the heights of the reference laser DTM and the predicted water bottom provides quantitative evidence of any systematic effects.

4.2 Reference Data Generation

4-band images (red, green, blue, coastal blue) serve as input data for deriving in-water slant ranges within our BathyNet CNN framework. The necessary steps are listed in the following:

- Fine geo-referencing of the laser bathymetry point clouds: We employed the rigorous strip adjustment approach described in Glira et al. (2016) as implemented in the scientific laser scanning software OPALS (Pfeifer et al. 2014). The RTK-GNSS surveyed control point targets served as the basis for co-registration of the concurrently captured laser and image data.
- Water surface estimation, Fig. 3a+c: To calculate a DWSM in regular grid structure (grid size: 10 m), we adopted the statistical approach of Mandlburger et al. (2013).
- Refraction correction: The raw 3D laser points need to be corrected due to laser ray refraction at the air–water interface and reduced propagation speed in water according to Snell's law. Refraction correction was carried out with OPALS using the laser beam vector (in 3D space) and a DWSM as basic input.
- ALB ground point filtering and DTM interpolation, Fig. 3b+c: Hierarchical robust interpolation (Pfeifer and Mandlburger 2018) was used to filter ground points above and below the water table. Thereafter, all ground points served as basis for interpolating a DTM in regular 0.5 m grid structure with linear prediction (Pfeifer and Mandlburger 2018).
- Water-Land Boundary (WLB), Fig. 3d: Intersection of DTM and DWSM yields the WLB-polygon as the zero-level contour line of the DTM-DWSM height difference model. The WLB is later used within the BathyNet framework to identify image patches containing either water areas only or both water and land.

- Image alignment via SfM, Fig. 3e: We employed Pix4D-Mapper (Strecha et al. 2012) for orienting the aerial images. Precise trajectory data based on GNSS and IMU (Inertial Measurement Unit) data served as basis for geo-referencing. In addition, RTK-GNSS surveyed ground control points (checker board targets) were available for two lakes (Autobahnsee, Friedberger See).
- Ray tracing, Fig. 3f: The intersection of the 3D image rays with the DWSM, subsequent ray direction correction due to refraction at the air–water interface, and intersection of the refracted ray with the DTM constitute the core part of the preprocessing pipeline. We first construct the 3D image ray vector using the exterior image orientation, the focal length, and the image pixel coordinates, and employ the Ebree library (Wald et al. 2014) for ray tracing. The main output of this processing step is the IWSR for each image pixel within the water domain representing the reference values for depth retrieval via BathyNet.

4.3 BathyNet

The core idea of our approach is to estimate depth information solely from airborne imagery. We opt for approximating a depth function D by employing a ML model (cf. Fig. 3g) in order to avoid fine tuning of the parameters of an analytical model (Lyzenga et al. 2006; Stumpf et al. 2003; Legleiter et al. 2009). Here, we distinguish between two core approaches, (i) feature-driven and (ii) data driven models. The first one relies on hand-crafted features like logarithms of certain spectral bands or band ratios (Makboul et al. 2017; Legleiter et al. 2009). Especially for 2D imagery such features are often hard to design, so that fitting a model directly to the data and learning the features automatically from the measurements is preferable (Sonogashira et al. 2020).

The most prominent representative of the latter are Convolutional Neural Networks, which are a more specific type of neural network that can be used to process images efficiently by taking advantage of their spatial structure (LeCun et al. 2015). This is realised by performing convolutions in the individual layers. Depending on the size of the kernel, small areas of e.g. 3×3 pixels from one layer provide information for a corresponding neuron in a deeper layer. By repeated pooling operations (i.e., sub-sampling) traversing deeper into the network, step by step a receptive field covering the initial image layer is established. Assuming homogeneity and translation invariance, effective feature extracting convolution kernels can be useful not only at one position, but anywhere in the image. Therefore, only one kernel is trained for the entire layer instead of training different kernels that refer to specific regions in the image. Thus, instead of training weights for each neuron in each layer, only the weights for one kernel need to be trained. This allows for deeper architectures with more layers to include high-level
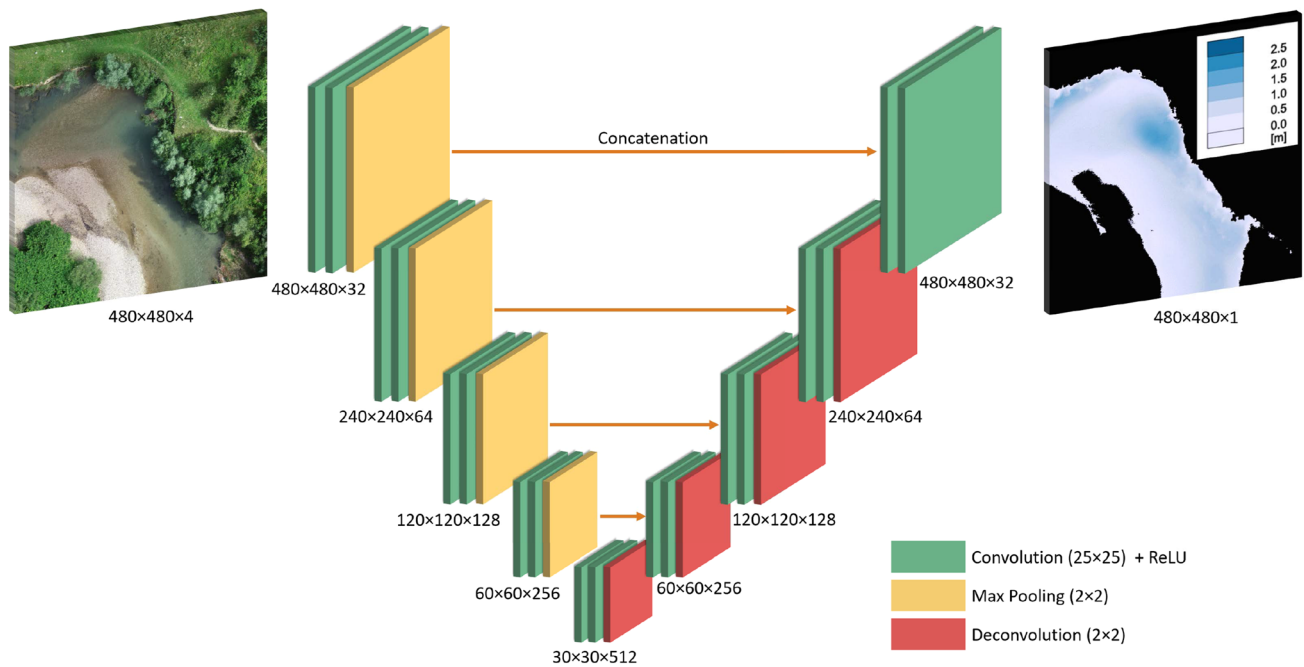
Fig. 4 Adapted U-Net architecture re-purposed for depth estimation

features, but still be efficient in terms of training time and required training data.

In the following, we focus on our specific architecture (Sect. 4.3.1) and on the specific adaption for our regression problem (Sect. 4.3.2).

4.3.1 Employed Architecture

As stated above, we intend to use our network for estimating continuous depth values for each individual pixel of a presented optical image. Therefore, an architecture where the input is both en- and decoded preserving the original shape of the input is required. For this, we rely on the state-of-the-art U-Net architecture as originally proposed by Ronneberger et al. (2015), which is commonly used for semantic segmentation of both 2D and 3D data (Rakhlin et al. 2018; Schmohl and Sörgel 2019).

The basic structure of U-Net is displayed in Fig. 4. Generally, U-Net is composed of an encoding and a decoding branch, which means that input images are first abstracted to a deep representation/understanding (which is located at the bottom of the "U"), where spatial context is mainly lost. In other words, we strive to find a mapping between a given high dimensional image space into a low dimensional representation such that we maintain the essence of the image content while spurious details and noise are suppressed. Afterwards, this deep representation is upsampled again to the original dimension but instead of the input channels, the prediction is obtained.

For encoding, each level composes two convolutional layers, followed by a Max Pooling layer for bisecting the input size, so that the maximum value of a respective sliding window is selected. This means that we set the receptive field to integrate the values of 4 pixels (filter size 2×2) in every second column and row (i.e., stride of 2, cf. Goodfellow et al. (2016)). This means that in each level of the encoding branch, we reduce the size of the input with the intention of allowing more convolutional filters (i.e. increase depth and allow a deeper understanding of the input images). More precisely, in the first level 32 filters per convolution are used, whereby the number of filters is doubled at each of our 4 levels to 512 filters at the end of the encoding branch, i.e., the deepest representation of our input image (cf. Fig. 4).

The subsequent upsampling (Dumoulin and Visin 2018) within the decoding branch is built in an analogous manner mainly differing in replacement of the Max Pooling layers by deconvolution layers, which are able to learn individual upconvolving filters and avoid the necessity of predefined functions such as nearest neighbour or bi-linear interpolation. While the encoder is able to efficiently encode the images, spatial context is lost due to the Max Pooling steps. Therefore, explicit connections between the encoding and decoding branch are built at each level. Precisely, the output of each respective encoding level together with the current state of representation is concatenated and used as input for each decoding level. For accelerating the whole learning process, batch normalisation is applied before each convolutional block. Following each convolution step, we use ReLU

as activation function allowing the approximation of a non-linear function.

Eventually, a 1×1 convolution is employed for mapping pixelwise feature vectors to the number of classes for semantic segmentation. As activation function typically the sigmoid function is used to turn the set of arbitrary output values into a-posteriori probabilities $p(c|x)$ that pixel x belongs to class c.

4.3.2 Reformulating U-Net for Water Depth Estimation as BathyNet

As stated above, we aim at using U-Net to estimate water depth. In contrast to the original formulation for semantic segmentation, we therefore need to adapt the architecture in order to allow pixelwise prediction of continuous depth values for our regression problem.

For this, we need to alter the last convolutional layer from feeding an output of the size of the input image times the number of classes (i.e. bands) to feeding only one band, which consists of the estimated IWSR values. Consequently, using a sigmoid as the activation function of this last convolution is not appropriate. In order to (i) allow continuous floating point values and (ii) avoid prediction of negative depth values (we interpret depth as positive floating point numbers), we make use of the ReLU activation function:

$$f(x) = \max(0, x) \tag{2}$$

Additionally, the root mean squared error (RMSE) is our objective function, for which we compare the depth value $\hat{y}_i$ predicted by the current state of our model to the reference value $\bar{y}_i$:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - \bar{y}_i)^2} \tag{3}$$

The kernel size of the convolutional layers is set to 25×25 pixels (empirically determined). For avoiding size reduction due to only partially filled filters at the border of the respective feature maps, we apply padding. Although CNNs incorporate by far fewer learnable parameters than NNs, we still need to provide informative training datasets. Overfitting might occur if insufficient data is provided for training due to the lack of examples of all possible real world scenarios. This means that the model learns overly specific characteristics from the provided data. In order to restrict this behaviour, we make use of Dropout layers at each U-Net level both for the encoding and decoding branch, where we randomly ignore a fixed percentage of neurons in each forward pass (25% of neurons in this case).

4.3.3 Data Preparation and Augmentation

Since we cannot directly train on a complete, full resolution image due to memory limitations, we generate image patches, i.e., we crop a limited area from the original image. The size of the patches needs to be divisible by $2^4 = 16$, in order to receive a natural number of pixels in width and height after the fourth and last pooling layer. We use a patch size of 480×480 pixels. This is a compromise between losing the complex characteristics of shore areas when using patches that are too small and minimizing memory consumption.

A certain number of these patches are extracted from each image of the training set. The patch count is set relative to the number of water pixels in the image, which we refer to as valid pixels. To be more precise, the number of patches is derived as ratio between pixels in the respective image and the filter size multiplied by an empirically derived factor of 2.7. For cropping these patches, we center our 480×480 pixel mask at a random position in the image. If less than 25% of pixels within this mask are valid pixels, we reject this patch. Otherwise we use it for training. This procedure ensures that, (i) each patch contains a reasonable number of valid pixels, (ii) no patches contain only land (i.e. invalid pixels), and (iii) shallow littoral areas are also present in the training set, which is crucial in order to cover a maximum range of depth values to be trained on.

Another option to enlarge the training data set is augmentation. Therefore, our training set is enhanced by flipping each patch generated around the x- and y-axis, transposing and rotation by 90 degrees both clockwise and counterclockwise.

4.4 3D Point Cloud Derivation

BathyNet yields depth (IWSR) predictions for every image pixel. In order to obtain 3D coordinates of the water bottom, generally the same steps as described in Sect. 4.2 are necessary including construction of the 3D image ray vector, intersection with the DWSM, and calculation of the direction of the refracted, subaqueous image ray. Thus, a water surface model needs to be available not only for training but also for depth retrieval via the trained deep learning model. For obtaining the final 3D bottom point coordinates, the subaqueous image ray unit vector is multiplied with the predicted IWSR and added to the position vector of the air-water intersection point.

Overlapping images result in redundant depth estimations from individual images at approximately the same *XY*-position. This redundancy can further be exploited for smoothing the final bottom model. For this, we aggregate bottom points from all images within 20 cm raster cells and calculate the final bottom point elevation of a certain raster cell as the

median elevation of all points in the cell (cf. Fig. 3h). Furthermore, we calculate the standard deviation σ_z of the cell point elevations which characterizes the precision of the bottom estimation. Wherever ground truth (i.e., a laser DTM) is available, we further calculate the bias dZ for each cell as:

$$dZ = z_{ref} - z_{BathyNet} \tag{4}$$

with z_{ref} denoting the laser DTM height and $z_{BathyNet}$ the estimated bottom height of a single raster cell derived with the described deep learning framework. We also calculate the mean error (ME) as the simple average of all dZ-values for each individual water body:

$$ME = \sum_i^n \frac{dZ_i}{n} \tag{5}$$

5 Results

In this section, we present the results of our BathyNet depth mapping approach. As detailed in Sect. 4.3, the employed U-Net CNN generally predicts in-water slant ranges for every pixel of an aerial image subject to prior model training with simultaneously captured laser bathymetry reference data. Thus, the initial results are in image space. In contrast, the final product is a continuous water depth model in object space, obtained by transforming the predicted IWSR values as outlined in Sect. 4.4. In the following, we assess the quality of the BathyNet derived in-water slant ranges (image space) and water depth maps (object space) by comparison to the respective laser reference data.

The implementation of BathyNet is realized in Python (Van Rossum and Drake 1995) using the deep learning library keras (Chollet et al. 2015) with tensorflow backend (Abadi et al. 2015). All computations were carried out on a desktop computer and the network was trained on a GeForce GTX 1070 GPU. When using all images of Helenensee, training of BathyNet takes about 156 minutes. Inference on individual images is carried out in less than 6 minutes. BathyNet is trained by usage of Adam optimizer. Learning rate was set to 0.00019 and training is conducted for 100 epochs.

Figure 5 summarizes the BathyNet results for Autobahnsee both in image and object space. Figure 5a shows a heat map of BathyNet-predicted IWSRs compared to the laser reference. The plot comprises the data of all available images used during training as well as testing. Most of the
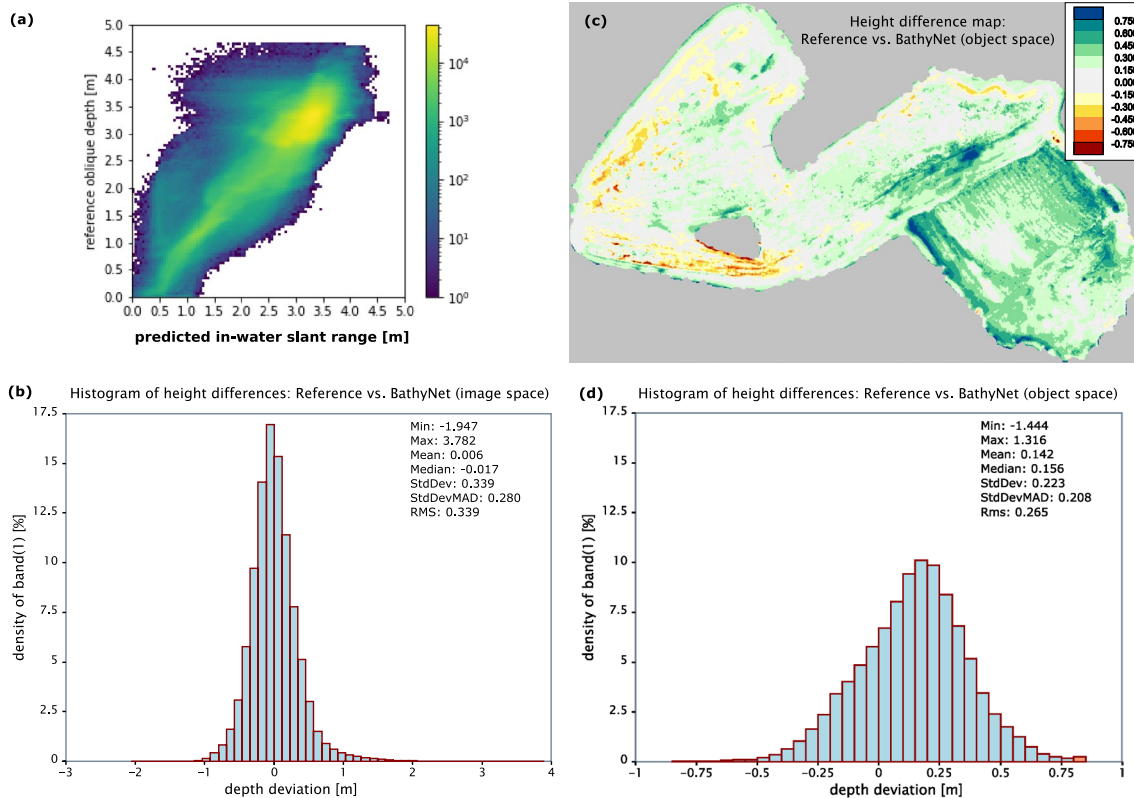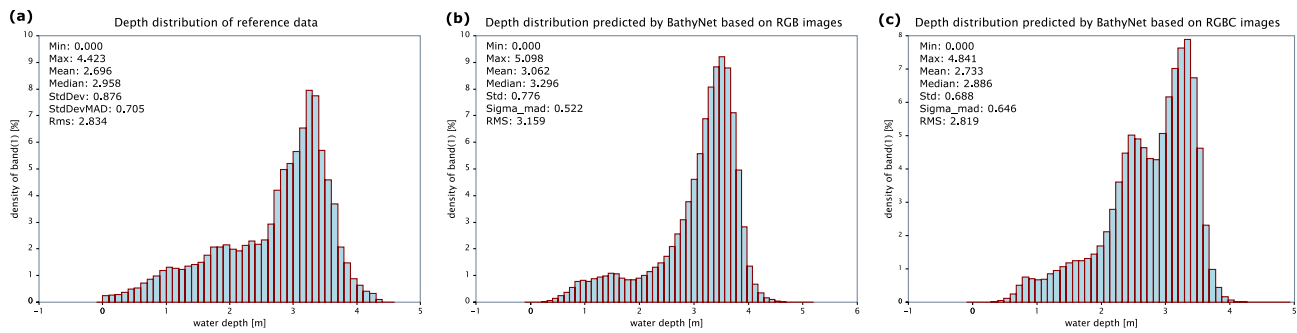

Fig. 5 BathyNet results for Autobahnsee; comparison of IWSR-values between laser reference and BathyNet predictions in image space as heat map (**a**) and histogram (**b**); nominal-actual water depth deviations in object space as colour coded DoD map (**c**) and histogram (**d**)

(a)

Depth distribution of reference data

Min: 0.000
Max: 4.423
Mean: 2.696
Median: 2.958
StdDev: 0.876
StdDevMAD: 0.705
Rms: 2.834

(b)

Depth distribution predicted by BathyNet based on RGB images

Min: 0.000
Max: 5.098
Mean: 3.062
Median: 3.296
Std: 0.776
Sigma_mad: 0.522
RMS: 3.159

(c)

Depth distribution predicted by BathyNet based on RGBC images

Min: 0.000
Max: 4.841
Mean: 2.733
Median: 2.886
Std: 0.688
Sigma_mad: 0.646
RMS: 2.819

Fig. 6 Depth distribution histograms; laser reference (**a**), BathyNet predictions based on RGB (**b**) and RGBC images (**c**)

values (yellow colour tones) are concentrated around the principal diagonal indicating a good coincidence between predicted and reference IWSR values. The histogram in Fig. 5b provides the corresponding quantitative analysis. The distribution is approximately unbiased (median: −2 cm) and the dispersion measures around 30 cm (conventionally estimated standard deviation StdDev: 34 cm, robustly estimated standard deviation StdDevMAD: 28 cm).

Figure 5c depicts a colour coded DEM of Difference (DoD) map, where each pixel contains the deviation between the reference depth obtained from the laser DTM and the BathyNet derived water bottom. The prevailing light green and orange colour tones indicate deviations of less than 20 cm. In contrast to the IWSR comparison, the depth deviation histogram (Fig. 5d) reveals a positive bias of 15 cm corresponding to the predominant greenish colour tones of the DoD map. Median filtering of the 3D point clouds derived from highly overlapping images, in turn, led to a decrease of the standard deviation to 20 cm, equivalent to roughly 5% of the maximum depth of 4.5 m.

The impact of using the additional Coastal Blue band is illustrated in the depth distribution histograms of Fig. 6, again exemplified for Autobahnsee. The laser reference depth distribution on the left side (Fig. 6a) is compared to those derived from BathyNet using RGB only (b) and including the additional Coastal Blue band (c), respectively. The depth model derived from RGBC images outperforms the RGB model in all statistical parameters. The RGB derived mean water depth (3.06 m), for instance, deviates from the reference value (2.70 m) by 26 cm, while the difference is only 3 cm when using RGBC images as input for BathyNet. Thus, the use of the Coastal Blue spectral band has a positive impact on depth retrieval in our shallow water study area.

The results shown so far only document the performance of BathyNet for data of the same water body. That is, training and testing relied on images of a single water body with consistent, uniform water quality and bottom texture. In the following, we therefore focus on the portability of trained BathyNet networks. For this, we trained BathyNet on one of the four lakes and inferred the depths of the remaining three lakes applying the trained model. Figure 7 summarizes the results for Autobahnsee (a), Friedberger See (b), and Kuhsee (c) trained with RGBC images of Helenensee.

The first two rows show the validation results in image space (IWSR comparison), and the latter three rows the respective results in object space (water depth maps). As in Fig. 5a and b, the first row of Fig. 7 depicts reference vs. BathyNet derived IWSR heat maps and the second row the corresponding histograms. Applying the model trained for Helenensee to Autobahnsee and Kuhsee yields acceptable results with a maximum mean offset of 15 cm and a standard deviation of 50 cm. The standard deviation is higher by a factor of 2.5 compared to the results achieved when performing training on the same water body (cf. Fig. 5b). For Friedberger See, both the heat map (b1) and the histogram (b2) reveal severe inaccuracies. The yellow (hot) areas of the heat map are not concentrated along the main axis diagonal as is the case for Autobahnsee (a1) and Kuhsee (c1). The mean depth deviation amounts to −1.07 m ± 1.18 m. Although the fact that Friedberger See is deeper compared to Autobahnsee and Kuhsee needs to be considered when judging the higher error values, these measures indicate a sub-optimal performance of BathyNet for this lake.

The water depth model validation shown in rows (3)–(5) of Fig. 7 generally exhibits the same tendency as the IWSR comparison. The systematic bias amounts to 12 cm for Autobahnsee and Kuhsee and the standard deviation is approximately 40 cm for both lakes. While training of BathyNet with images from Helenensee does not increase the systematic depth biases for Autobahnsee and Kuhsee, the standard deviation increases by a factor of 2. For Autobahnsee, the increased spread measures approximately 10% of the maximum depth. The same holds for Friedberger See with a mean water depth underestimation of 1 m and a dispersion of 75 cm.

The last row of Fig. 7 shows a colour coded map of local depth deviations derived from the 3D point cloud of

overlapping images. Each pixel shows the standard deviation of the point heights within a 20 cm raster cell. This measure is an indicator for the locally varying water depth uncertainty. The predominant green and blue colour tones for Autobahnsee (a5) and Kuhsee (c5), denoting a dispersion less than 30 cm, confirm a good repeatability of the BathyNet depth retrieval from independent images with varying viewing geometry (image overlap: 90%). Only the deeper, northwestern part of Kuhsee reveals standard deviations greater than 50 cm. In contrast, saturated red colour tones corresponding to standard deviations ≥ 1 m dominate the dispersion map of Friedberger See (b5), highlighting once again that the network trained on Helenensee is not directly applicable to Friedberger See due to differences in (i) water depth, (ii) water quality, and (iii) bottom reflectance.

6 Discussion

In this section, we first discuss the results reported in Sect. 5 including statements concerning the limitations of the presented framework in its present form. Based on the drawn conclusions, we anticipate an extension of BathyNet for simultaneous depth retrieval and classification of bare water bottom and submerged vegetation.

6.1 General Discussion of Results

Spectrally based depth retrieval works best in areas with homogeneous bottom texture whereas geometry based methods like multimedia photogrammetry inherently require texture, i.e. spatial variations in the spectral response (Slocum et al. 2020). While it is possible to reduce the depth estimation errors to a certain extent by means of multispectral algorithms (Lyzenga et al. 2006; Legleiter and Fosness 2019), reflectance variations still pose a fundamental problem, especially in the very shallow littoral zone where submerged vegetation is the main source of spectral differences rather than the homogeneity of the water bottom substrate (Heblinski et al. 2011). In addition, the existence of underwater vegetation is also problematic for the generation of reference data, for which echo sounding or laser bathymetry are the most prominent capturing techniques (Lyzenga et al. 2006; Song et al. 2015; Kasvi et al. 2019; Rossi et al. 2020; Brown et al. 2011; Kogut and Bakuła 2019). Both methods are generally capable of penetrating (loose) vegetation and therefore often provide reference depths related to the bottom whereas image based techniques tend to deliver the topmost surface of the vegetation canopy (Ressl et al. 2016).

All of the groundwater supplied lakes in this study comprise both bare gravel areas as well as littoral vegetation featuring variable brightness and vegetation height. While some species cover the ground and thereby change the bottom reflectance, other species tend to reach the water surface in the spring season and therefore substantially extend above the water bottom. Both effects hamper SDB methods in general and also limit the accuracy of our BathyNet approach in particular. Still, the achieved error values are satisfactory, especially considering the aforementioned external influences. BathyNet derived water depth models showed a systematic deviation (mean error) in the range of 0–15 cm and a standard deviation of 20–40 cm after median filtering of the point clouds from overlapping images. This is equivalent to approximately 5% of the maximum depth. Compared to other published SDB approaches, BathyNet performs reasonably well. Kasvi et al. (2019) report ME and standard deviations in the dm-range for clear gravel bed rivers with homogeneous bottom texture. In contrast, the RMSE values in Hernandez and Armstrong (2016) measure 1–2 m for a band ratio SDB approach based on WorldView-2 images of a scene with heterogeneous bottom reflectance (seagrass etc.). Thus, when comparing the reported error metrics of different publications, all influencing factors like bottom homogeneity, water transparency, etc. need to be considered for judging the results.

A further investigation addressed the question whether the Coastal Blue band provides added value for our BathyNet depth retrieval. In prior work (Mandlburger et al. 2018), we have disclosed the transmission curves of the used sensor and also provided exemplary images revealing more pronounced bottom features in the Coastal Blue band ($\lambda \approx 430$ nm) compared to the regular blue band ($\lambda \approx 480$ nm) in some areas, although water penetrability generally decreases towards the ultraviolet domain of the spectrum. The quantitative comparison of two model variants, one with RGB only and the other with the additional Coastal Blue channel, clearly confirmed a positive impact of this extra radiometric channel in our study. The deep neural network was better able to approximate the laser reference data, and the results clearly outperformed the RGB-only model in every aspect of the statistical depth distribution analysis.

Besides the pure accuracy of the method (i.e., its ability to predict water depths based on spectral information as input), the portability of trained models to other water bodies was another point of focus. With three rather similar lakes concerning bottom reflectance, water transparency, and depth distribution (Autobahnsee, Helenensee, Kuhsee; cf. Fig. 1) and one lake with different characteristics (Friedberger See), training BathyNet with one of the prior lakes and predicting the depths of all remaining lakes exhibits mainly consistent results, which can be summarized as follows:

- Training the network with a subset of the images and testing the trained model with the remaining images of the
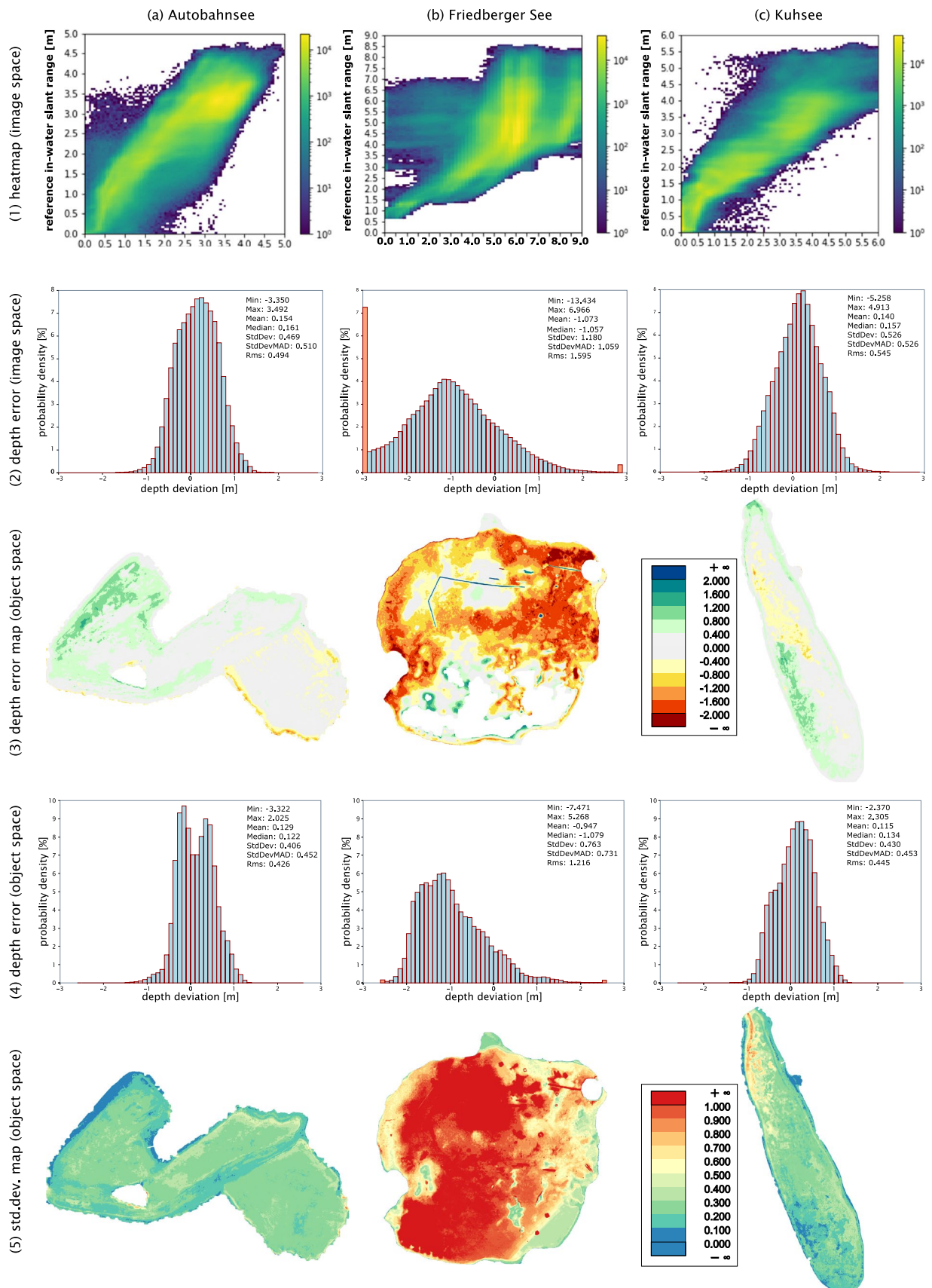
◀**Fig. 7** Transfer of BathyNet results trained on Helensee to Autobahnsee (**a**), Friedberger See (**b**), and Kuhsee (**c**); rows (1)+(2): IWSR comparison in image space; rows (3)–(5): water depth model comparison in object space; (1) heat maps of reference vs. BathyNet-derived in-water slant ranges; (2) corresponding depth deviation histogram; row (3): colour coded nominal-actual water depth DoD map; row (4): corresponding depth deviation histogram; row (5): colour coded standard deviation map

same lake gives the best results. This scenario is relevant for mapping large extended water bodies, where reference data is available for a small portion of the area.

- Inference of water depths using a model trained on a different lake with similar characteristics does not affect the mean error but increases the standard deviation by a factor of around 2.
- The latter behaviour was observed for depth predictions of Autobahnsee and Kuhsee with a model trained on Helenensee. The results are less favourable when training on Autobahnsee and inferring depths of Helenensee.
- Helenensee (mean/max depth: 3.5 m/7.0 m) covers a larger depth range compared to Autobahnsee (mean/max depth: 2.9 m/4.5 m) and Kuhsee (mean/max depth: 1.5 m/5.3 m). Thus, training the network on Helenensee results in interpolation when inferring the depths of Autobahnsee and Kuhsee, whereas this would lead to depth extrapolation for Helenensee. The importance of representative training data was also highlighted by Legleiter and Fosness (2019).
- Regardless which of the three lakes was used for training, the predicted depths for Friedberger See were always sub-optimal. Friedberger See is the deepest of the four studied lakes with a maximum depth of around 12 m beyond the laser penetration depth of 8 m, and the lake contains large continuous underwater vegetation areas and an overall lower percentage of bare gravel bottom. Furthermore, the lake is used for recreational purposes with many artificial objects (floating platforms, infrastructure for water skiing, etc.), which also hampers spectrally based depth estimation. Therefore, this lake was not used for training but only for inference.
- Despite the unfavourable accuracy, the Friedberger See test site reveals a benefit of SDB in general and BathyNet in particular, namely the possibility to obtain a gapless depth map. Although parts of the lake exceeded the maximum penetration depth of the laser, it is still possible to derive a realistic depth model with restricted accuracy due to extrapolation.
- Our approach combines strict geometric processing of the generally oblique image rays including refraction correction at the water surface and radiometry based depth retrieval. This implies that training as well as depth inference requires a continuous water surface model. While the DSWM could be derived from the laser measure-

ments in our study, availability of laser bathymetry data is not given for image-only surveys. In this case, the DSWM can for instance be reconstructed from the multispectral images (Frazier and Page 2000; Wang et al. 2020) or from tidal data.

Although BathyNet is capable of predicting continuous water depth values, we conclude that the main factor limiting the accuracy of our BathyNet is the abundant occurrence of littoral vegetation. Seasonal changes of underwater vegetation and water transparency pose another issue in this context. The basic idea to circumvent these limitations is outlined in the following.

6.2 Expansion of BathyNet

The BathyNet framework described in Sect. 4 does not distinguish between bare underwater soil and aquatic vegetation. This means that BathyNet approximates the same function D for all input data. However, as the impact of bottom reflectance variations on depth errors can only be mitigated using multispectral algorithms (Lyzenga et al. 2006), we aim at combining underwater land cover classification and depth estimation in a comprehensive deep learning framework. In particular, our goal is to enhance BathyNet in such a way that a vegetation mask M is predicted on top of depth values. The prior constitutes a binary semantic segmentation, for which U-Net was originally designed (cf. Sect. 4.3.1).

Although training two U-Nets, one for semantic segmentation deriving M and one for the regression approximating D, is straightforward, we would have to train twice as much weights as before. Furthermore, since we would feed the exact same input to the two nets, we opt for incorporating both interpretation tasks in one joint U-Net, so that weights can be shared for both tasks. Precisely, we only add a second output branch (1×1 convolutional layer) eventually obtaining two different outputs, namely (i) pixelwise depth values (shape: input dimension $\times 1$) and (ii) a two-class semantic segmentation result (shape: input dimension $\times 2$ classes).

To demonstrate the general feasibility, Fig. 8 shows a raw image (left), the corresponding ground truth vegetation polygons provided by limnologists (middle), and the derived BathyNet vegetation mask (right). Further development of the combined model and proper quantitative error assessment of both bathymetry and binary classification is beyond the scope of this paper and subject of future work.

7 Conclusions

In this article, we presented a comprehensive method for water depth retrieval from multispectral aerial images. Our BathyNet approach combines rigorous geometric image
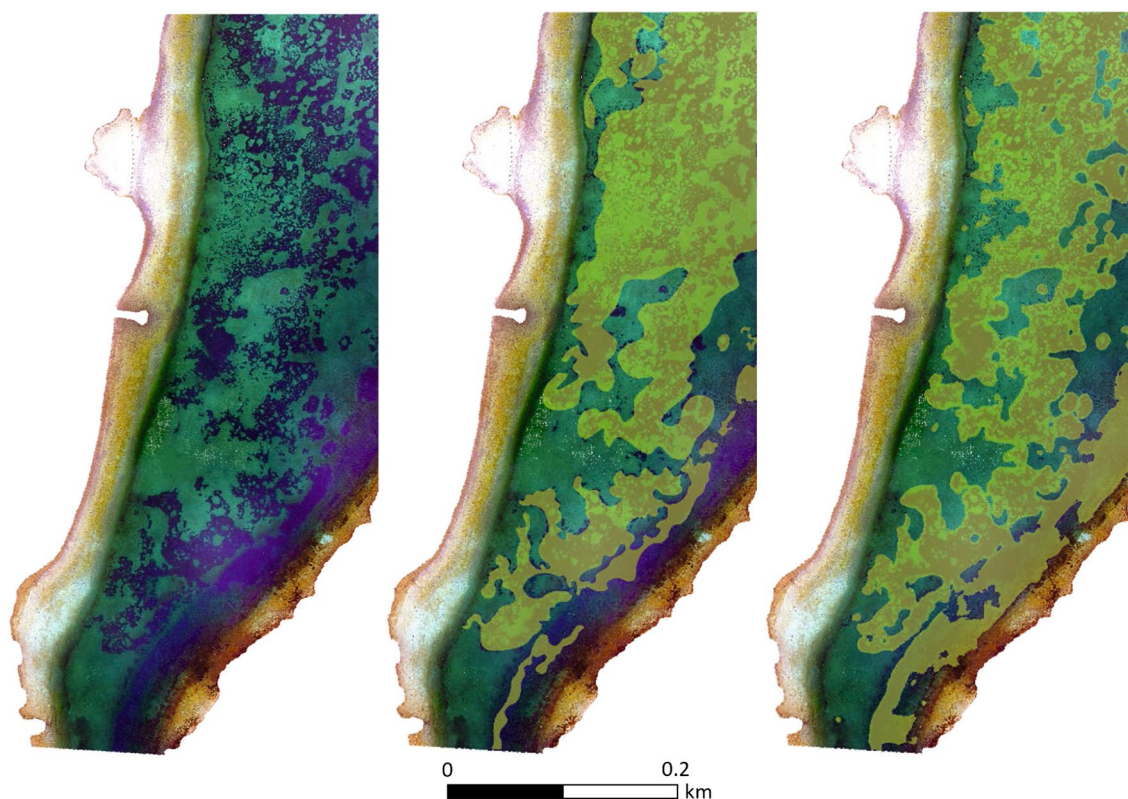
Fig. 8 Comparison of a raw image (left), ground truth vegetation mask (middle), and predicted vegetation mask (right) for the Kuhsee test site

ray path modelling, including refraction correction at the air–water interface, and spectrally based estimation of in-water slant ranges (IWSR) within a deep learning framework. For the latter, we adapted the U-Net Convolutional Neural Network architecture, originally designed for semantic segmentation, for the purpose of pixelwise depth estimation. Thus, BathyNet connects rigorous photogrammetric treatment of the image geometry and state-of-the art CNN-based processing of the multispectral image content.

Neural networks generally require a vast amount of training data for estimating the model parameters (weights of filter kernels, etc.). In our study, we employed concurrently captured laser bathymetry data for generating reference data. Due to the increasing availability of hybrid sensor systems comprising both laser scanners and multispectral cameras, we believe that such data combinations will gain importance in the future. Data preprocessing involved the generation of a continuous water surface model needed for proper refraction correction of the laser and image rays, as well as the interpolation of an underwater DTM (i.e., water bottom model). Both models served as the basis for the calculation of IWSRs for every image pixel. Together with the Red, Green, Blue, and Coastal Blue (RGBC) spectral bands, the slant ranges constituted the input for training of our BathyNet model. After training, the model was capable of predicting IWSRs

on a per pixel basis. A smooth water depth model was finally achieved by transforming the predicted slant ranges into object space and median filtering of the resulting 3D point clouds from overlapping images.

In a real-world application scenario, where multispectral images but no laser data are captured, the necessity of a DWSM for depth estimation seems a specific drawback of our method. However, our approach does not only deliver relative water depths at the time of data acquisition, but absolute 3D bottom topography in a specific coordinate reference system. From this point of view, any SDB method aiming at 3D reconstruction of underwater topography requires a model of the water surface. Assuming a horizontal water surface, it is feasible to derive the water level height at the time of data acquisition from the images by identifying the water-land boundary either manually or automatically. Answering the question how deviations from such a simplified water surface model impact the results of BathyNet is beyond the scope of this article and subject of future work.

We tested our approach with RGBC images of four groundwater supplied lakes located around Augsburg, Germany. The lakes feature a mean depth of 2–5 m and a maximum depth of 4.5–12 m. Training of BathyNet on a representative lake and inferring the depths from unseen images of the other lakes resulted in a mean systematic offset of 15 cm

and a standard deviation of approximately 40 cm, corresponding to 5–10% of the maximum water depth. These error values meet the vertical accuracy limit Bc7 defined by the International Hydrographic Organization (IHO 2020), but not to the more rigorous standards (Order 1, Special Order, Exclusive Order). While the use of the additional Coastal Blue band had a clear positive impact on the consistency of the predicted depths, the frequent change of bottom reflectance from bright gravel to dark littoral vegetation limited the accuracy. This observation was the starting point for an envisaged extension of BathyNet from a pure depth retrieval framework to a comprehensive method featuring simultaneous semantic segmentation (ground/vegetation) and bathymetry estimation. While the general feasibility of such an approach could already be verified in preliminary tests, further refinement of the method including proper quality assessment is subject of future work.

Declarations

References

Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X (2015) TensorFlow: large-scale machine learning on heterogeneous systems. http://tensorflow.org/. Software available from tensorflow.org

Agrafiotis P, Skarlatos D, Georgopoulos A, Karantzalos K (2019) DepthLearn: Learning to Correct the Refraction on Point Clouds Derived from Aerial Imagery for Accurate Dense Shallow Water Bathymetry Based on SVMs-Fusion with LiDAR Point Clouds. Remote Sensing 11(19): . https://doi.org/10.3390/rs11192225. https://www.mdpi.com/2072-4292/11/19/2225

Agrafiotis P, Karantzalos K, Georgopoulos A, Skarlatos D (2020) Correcting image refraction: towards accurate aerial image-based bathymetry mapping in shallow waters. Remote Sens. https://doi.org/10.3390/rs12020322. https://www.mdpi.com/2072-4292/12/2/322

Bergsma EWJ, Conley DC, Davidson MA, Hare TJ, Almar R (2019) Storm event to seasonal evolution of nearshore bathymetry derived from shore-based video imagery. Remote Sens. https://doi.org/10.3390/rs11050519.https://www.mdpi.com/2072-4292/11/5/519

Brown CJ, Smith SJ, Lawton P, Anderson JT (2011) Benthic habitat mapping: a review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuar Coast Shelf Sci 92(3):502–520. https://doi.org/10.1016/j.ecss.2011.02.007

Bué I, Catalão J, Semedo Á (2020) Intertidal bathymetry extraction with multispectral images: a logistic regression approach. Remote Sens. https://doi.org/10.3390/rs12081311. https://www.mdpi.com/2072-4292/12/8/1311

Cahalane C, Magee A, Monteys X, Casal G, Hanafin J, Harris P (2019) A comparison of Landsat 8, RapidEye and Pleiades products for improving empirical predictions of satellite-derived bathymetry. Remote Sens Environ 233: https://doi.org/10.1016/j.rse.2019.111414

Carrivick JL, Smith MW (2019) Fluvial and aquatic applications of Structure from Motion photogrammetry and unmanned aerial vehicle/drone technology. WIREs Water 6(1):e1328. https://doi.org/10.1002/wat2.1328. https://onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1328

Chollet F et al (2015) Keras. https://github.com/fchollet/keras

Dickens K, Armstrong A (2019) Application of machine learning in satellite derived bathymetry and coastline detection. SMU Data Sci Rev 2(1)

Dietrich JT (2016) Bathymetric Structure-from-Motion: extracting shallow stream bathymetry from multi-view stereo photogrammetry. Earth Surf Process Landf 42(2):355–364. https://doi.org/10.1002/esp.4060. https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.4060

Dumoulin V, Visin F (2018) A guide to convolution arithmetic for deep learning. eprint:1603.07285

Duplančić Leder T, Leder N, Peroš J (2019) Satellite derived bathymetry survey method—example of Hramina Bay. ACM Trans Math Softw 8:99–108

Effler SW (1988) Secchi disc transparency and turbidity. J Environ Eng 114(6):1436–1447. https://doi.org/10.1061/(ASCE)0733-9372(1988)114:6(1436)

Eugenio F, Marcello J, Martin J, Rodríguez-Esparragón D (2017) Benthic habitat mapping using multispectral high-resolution imagery: evaluation of shallow water atmospheric correction techniques. Sensors. https://doi.org/10.3390/s17112639

European Union (1992) Council Directive 92/43/EEC on the Conservation of natural habitats and of wild fauna and flora. Off J Eur Commun (OJL) L 206(35)

European Union (2000) Directive 2000/60/EC of the European Parliament and of the council of 23 October 2000 establishing a framework for Community action the field of water policy. Off J Eur Commun (OJL) L 327(173): 1–72

European Union (2007) Directive 2007/60/EC of the European Parliament and European Council of October 2007 on the assessment and management of flood risks . Off J Eur Commun (OJL) 288(27)

Frazier PS, Page KJ (2000) Water body detection and delineation with Landsat TM data. PE&RS Photogramm Eng Remote Sens 66(12):1461–1467

Gao J (2009) Bathymetric mapping by means of remote sensing: methods, accuracy and limitations. Progress Phys Geogr Earth Environ 33(1):103–116. https://doi.org/10.1177/0309133309105657

Gentile V, Mróz M, Spitoni M, Lejot J, Piógay H, Demarchi L (2016) Bathymetric mapping of shallow rivers with UAV hyperspectral data. In: Proceedings of the fifth international conference on telecommunications and remote sensing—volume 1: ICTRS, pp 43–49. INSTICC, SciTePress. https://doi.org/10.5220/0006227000430049

Glira P, Pfeifer N, Mandlburger G (2016) Rigorous strip adjustment of UAV-based laserscanning data including time-dependent correction of trajectory errors. Photogramm Eng Remote Sens 82(12):945–954. https://doi.org/10.14358/PERS.82.12.945. Special Issue on Mobile Mapping Technologies

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. Adaptive computation and machine learning. MIT Press, Cambridge. http://www.deeplearningbook.org

Guenther GC, Cunningham AG, Laroque PE, Reid DJ (2000) Meeting the accuracy challenge in airborne lidar bathymetry. In: Proceedings of the 20th EARSeL symposium: workshop on lidar remote sensing of land and sea. Dresden, Germany

Heblinski J, Schmieder K, Heege T, Agyemang TK, Sayadyan H, Vardanyan L (2011) High-resolution satellite remote sensing of littoral vegetation of lake sevan (armenia) as a basis for monitoring and assessment. Hydrobiologia 661:97–111. https://doi.org/10.1007/s10750-010-0466-6

Hernandez WJ, Armstrong RA (2016) Deriving bathymetry from multispectral remote sensing data. J Mar Sci Eng. https://doi.org/10.3390/jmse4010008. https://www.mdpi.com/2077-1312/4/1/8

Hilldale R, Raff D (2008) Assessing the ability of airborne LiDAR to map river bathymetry. Earth Surf Process Landf 33(5):773–783. https://doi.org/10.1002/esp.1575

Hodúl M, Bird S, Knudby A, Chénier R (2018) Satellite derived photogrammetric bathymetry. ISPRS J Photogramm Remote Sens 142:268–277. https://doi.org/10.1016/j.isprsjprs.2018.06.015

IHO: S-44 (2020) Standards for Hydrographic Surveys. Standard 6th ed., International Hydrographic Organization, Monaco. https://iho.int/uploads/user/pubs/Drafts/S-44_Edition_6.0.0-Final.pdf

Kasvi E, Salmela J, Lotsari E, Kumpula T, Lane S (2019) Comparison of remote sensing based approaches for mapping bathymetry of shallow, clear water rivers. Geomorphology 333:180–197. https://doi.org/10.1016/j.geomorph.2019.02.017

Kogut T, Bakuła (2019) Improvement of full waveform airborne laser bathymetry data processing based on waves of neighborhood points. Remote Sens. https://doi.org/10.3390/rs11101255. https://www.mdpi.com/2072-4292/11/10/1255

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

Legleiter CJ, Fosness RL (2019) Defining the limits of spectrally based bathymetric mapping on a large river. Remote Sens. https://doi.org/10.3390/rs11060665. https://www.mdpi.com/2072-4292/11/6/665

Legleiter CJ, Harrison LR (2019) Remote sensing of river bathymetry: evaluating a range of sensors, platforms, and algorithms on the Upper Sacramento River, California, USA. Water Resour Res 55(3):142–2169. https://doi.org/10.1029/2018WR023586. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023586

Legleiter CJ, Roberts DA, Lawrence RL (2009) Spectrally based remote sensing of river bathymetry. Earth Surf Process Landf 34(8):1039–1059. https://doi.org/10.1002/esp.1787

Li R, Liu W, Yang L, Sun S, Hu W, Zhang F, Li W (2018) DeepUNet: a deep fully convolutional network for pixel-level sea-land segmentation. IEEE J Select Top Appl Earth Observ Remote Sens 11(11):3954–3962. https://doi.org/10.1109/JSTARS.2018.2833382

Liu S, Wang L, Liu H, Su H, Li X, Zheng W (2018) Deriving bathymetry from optical images with a localized neural network algorithm. IEEE Trans Geosci Remote Sens 56(9):5334–5342. https://doi.org/10.1109/TGRS.2018.2814012

Lyzenga DR (1978) Passive remote sensing techniques for mapping water depth and bottom features. Appl Opt 17(3):379–383. https://doi.org/10.1364/AO.17.000379

Lyzenga DR, Malinas NP, Tanis FJ (2006) Multispectral bathymetry using a simple physically based algorithm. IEEE Trans Geosci Remote Sens 44(8):2251–2259. https://doi.org/10.1109/TGRS.2006.872909

Ma Y, Xu N, Liu Z, Yang B, Yang F, Wang XH, Li S (2020) Satellite-derived bathymetry using the ICESat-2 lidar and Sentinel-2 imagery datasets. Remote Sens Environ 250: https://doi.org/10.1016/j.rse.2020.112047

Maas HG (2015) On the accuracy potential in underwater/multimedia photogrammetry. Sensors 15(8):18140–18152. https://doi.org/10.3390/s150818140. http://www.mdpi.com/1424-8220/15/8/18140

Makboul O, Negm A, Mesbah S, Mohasseb M (2017) Performance assessment of ANN in estimating remotely sensed extracted bathymetry. Case study: eastern harbor of Alexandria. Procedia Eng 181:912–919. 10th International Conference Interdisciplinarity in Engineering, INTER-ENG (2016) 6–7 October 2016, Tirgu Mures. Romania. https://doi.org/10.1016/j.proeng.2017.02.486

Mandlburger G (2019) Through-water dense image matching for shallow water bathymetry. Photogramm Eng Remote Sens. https://doi.org/10.14358/PERS.85.6.445

Mandlburger G (2020) A review of airborne laser bathymetry for mapping of inland and coastal waters. J Appl Hydrogr 116:6–15. https://doi.org/10.23784/HN116-01

Mandlburger G, Pfennigbauer M, Pfeifer N (2013) Analyzing near water surface penetration in laser bathymetry—a case study at the River Pielach. In: ISPRS annals of the photogrammetry, remote sensing and spatial information sciences, vol 2. https://doi.org/10.5194/isprsannals-II-5-W2-175-2013

Mandlburger G, Hauer C, Wieser M, Pfeifer N (2015) Topo-bathymetric LiDAR for monitoring river morphodynamics and instream habitats—a case study at the Pielach River. Remote Sens 7(5):6160–6195. https://doi.org/10.3390/rs70506160. http://www.mdpi.com/2072-4292/7/5/6160

Mandlburger G, Kremer J, Steinbacher F, Baran R (2018) Investigating the use of coastal blue imagery for bathymetric mapping of inland water bodies. In: International archives of the photogrammetry, remote sensing and spatial information sciences—ISPRS Archives, 1/42. https://doi.org/10.5194/isprs-archives-XLII-1-275-2018

Misra A, Ramakrishnan B (2020) Assessment of coastal geomorphological changes using multi-temporal satellite-derived bathymetry. Continent Shelf Res 207: https://doi.org/10.1016/j.csr.2020.104213

Misra A, Vojinovic Z, Ramakrishnan B, Luijendijk A, Ranasinghe R (2018) Shallow water bathymetry mapping using Support Vector Machine (SVM) technique and multispectral imagery. Int J Remote Sens 39(13):4431–4450. https://doi.org/10.1080/01431161.2017.1421796

Muzirafuti A, Barreca G, Crupi A, Faina G, Paltrinieri D, Lanza S, Randazzo G (2020) The contribution of multispectral satellite image to shallow water bathymetry mapping on the Coast of Misano Adriatico, Italy. J Mar Sci Eng. https://doi.org/10.3390/jmse8020126. https://www.mdpi.com/2077-1312/8/2/126

Parrish CE, Dijkstra JA, O'Neil-Dunne JPM, McKenna L, Pe'eri S (2016) Post-Sandy Benthic Habitat Mapping Using New Topobathymetric Lidar Technology and Object-Based Image

Classification. Journal of Coastal Research pp 200–208. https://doi.org/10.2112/SI76-017

Parrish CE, Magruder LA, Neuenschwander AL, Forfinski-Sarkozi N, Alonzo M, Jasinski M (2019) Validation of ICESat-2 ATLAS bathymetry and analysis of ATLAS's bathymetric mapping performance. Remote Sens. https://doi.org/10.3390/rs11141634. https://www.mdpi.com/2072-4292/11/14/1634

Pfeifer N, Mandlburger G (2018) Lidar data filtering and Digital Terrain Model generation. In: Shan J, Toth CK (eds) Topographic laser ranging and scanning—principles and processing, 2nd edn. CRC Press, Boca Raton, pp 349–378

Pfeifer N, Mandlburger G, Otepka J, Karel W (2014) OPALS—a framework for Airborne Laser Scanning data analysis. Comput Environ Urban Syst 45:125–136. https://doi.org/10.1016/j.compenvurbsys.2013.11.002. http://www.sciencedirect.com/science/article/pii/S0198971513001051

Philpot W (2019) Airborne Laser Hydrography II. Cornell University Library (eCommons), Coernell. https://doi.org/10.7298/jxm9-g971. https://ecommons.cornell.edu/handle/1813/66666

Purkis SJ, Gleason ACR, Purkis CR, Dempsey AC, Renaud PG, Faisal M, Saul S, Kerr JM (2019) High-resolution habitat and bathymetry maps for 65,000 sq km of Earth's remotest coral reefs. Coral Reefs 38(3):467–488. https://doi.org/10.1007/s00338-019-01802-y

Rakhlin A, Davydow A, Nikolenko S (2018) Land cover classification from satellite imagery with U-Net and Lovász-Softmax loss. In; 2018 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW), pp 257–2574

Rasquin C, Seiffert R, Wachler B, Winkel N (2020) The significance of coastal bathymetry representation for modelling the tidal response to mean sea level rise in the german bight. Ocean Sci 16(1):31–44. https://doi.org/10.5194/os-16-31-2020. https://os.copernicus.org/articles/16/31/2020/

Ressl C, Brockmann H, Mandlburger G (2016) Pfeifer N (2016) Dense Image matching vs airborne laser scanning–comparison of two methods for deriving terrain models. Photogrammetrie Fernerkundung Geoinf 2:57–73. https://doi.org/10.1127/pfg/2016/0288

Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds) Medical image computing and computer-assisted intervention—MICCAI 2015. Springer, Cham, pp 234–241

Rossi L, Mammi I, Pelliccia F (2020) UAV-derived multispectral bathymetry. Remote Sens. https://doi.org/10.3390/rs12233897. https://www.mdpi.com/2072-4292/12/23/3897

Sagawa T, Yamashita Y, Okumura T, Yamanokuchi T (2019) Satellite derived bathymetry using machine learning and multi-temporal satellite images. Remote Sens. https://doi.org/10.3390/rs11101155. https://www.mdpi.com/2072-4292/11/10/1155

Salameh E, Frappart F, Almar R, Baptista P, Heygster G, Lubac B, Raucoules D, Almeida LP, Bergsma EWJ, Capo S, De Michele M, Idier D, Li Z, Marieu V, Poupardin A, Silva PA, Turki I, Laignel B (2019) Monitoring beach topography and nearshore bathymetry using spaceborne remote sensing: a review. Remote Sens. https://doi.org/10.3390/rs11192212. https://www.mdpi.com/2072-4292/11/19/2212

Sam L, Prusty G, Gahlot N (2018) Evaluation of optical remote sensing-based shallow water bathymetry for recursive mapping. Geocarto Int 33(7):737–753. https://doi.org/10.1080/10106049.2017.1299800

Schmohl S, Sörgel U (2019) Submanifold sparse convolutional networks for semantic segmentation of large-scale ALS point clouds. ISPRS Ann IV-2/W5:77–84. https://doi.org/10.5194/isprs-annals-IV-2-W5-77-2019. https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2-W5/77/2019/

Slocum RK, Parrish CE, Simpson CH (2020) Combined geometric-radiometric and neural network approach to shallow bathymetric mapping with uas imagery. ISPRS J Photogramm Remote Sens 169:351–363. https://doi.org/10.1016/j.isprsjprs.2020.09.002

Song Y, Niemeyer J, Ellmer W, Soergel U, Heipke C (2015) Comparison of three airborne laser bathymetry data sets for monitoring the German Baltic Sea Coast. https://doi.org/10.1117/12.2194960

Sonogashira M, Shonai M, Iiyama M (2020) High-resolution bathymetry by deep-learning-based image superresolution. PLOS One 15(7):1–19. https://doi.org/10.1371/journal.pone.0235487

Starek MJ, Giessel J (2017) Fusion of uas-based structure-from-motion and optical inversion for seamless topo-bathymetric mapping. In: 2017 IEEE international geoscience and remote sensing symposium (IGARSS), pp 2999–3002. https://doi.org/10.1109/IGARSS.2017.8127629

Strecha C, Küng O, Fua P (2012) Automatic mapping from ultra-light uav imagery. In: EuroCOW, 2012, Barcelona, Spain. http://infoscience.epfl.ch/record/175351

Stumpf RP, Holderied K, Sinclair M (2003) Determination of water depth with high-resolution satellite imagery over variable bottom types. Limnol Oceanogr 48(1part2):547–556

Szeliski R (2011) Computer vision: algorithms and applications. Springer, London. https://doi.org/10.1007/978-1-84882-935-0

Van Rossum G, Drake Jr FL (1995) Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam

Wald I, Woop S, Benthin C, Johnson GS, Ernst M (2014) Embree: a kernel framework for efficient CPU ray tracing. ACM Trans Gr. doi 10(1145/2601097):2601199

Wang G, Wu M, Wei X, Song H (2020) Water Identification from High-Resolution Remote Sensing Images Based on Multidimensional Densely Connected Convolutional Neural Networks. Remote Sens. https://doi.org/10.3390/rs12050795. https://www.mdpi.com/2072-4292/12/5/795

Westaway RM, Lane SN, Hicks DM (2001) Remote sensing of clear-water, shallow, gravel-bed rivers using digital photogrammetry. Photogramm Eng Remote Sens 67(11):1271–1281

Westfeld P, Maas HG, Richter K, Weiß R (2017) Analysis and correction of ocean wave pattern induced systematic coordinate errors in airborne LiDAR bathymetry. ISPRS J Photogramm Remote Sens 128:314–325. https://doi.org/10.1016/j.isprsjprs.2017.04.008

Zhang X, Ma Y, Zhang J (2020) Shallow water bathymetry based on inherent optical properties using high spatial resolution multispectral imagery. Remote Sens. https://doi.org/10.3390/rs12183027

Zhou DX (2020) Universality of deep convolutional neural networks. Appl Comput Harmon Anal 48(2):787–794. https://doi.org/10.1016/j.acha.2019.06.004