https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

Atmospheric Measurement Techniques
Discussions

Deep-Pathfinder: A boundary layer height detection algorithm based on image segmentation

Jasper S. Wijnands[1], Arnoud Apituley[1], Diego Alves Gouveia[1], and Jan Willem Noteboom[1]

[1]Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands

Correspondence: Jasper S. Wijnands (jasper.wijnands@knmi.nl)

Abstract. The mixing layer height (MLH) indicates the change between vertical mixing of air near the surface and less turbulent air above. MLH is important for the dispersion of air pollutants and greenhouse gases, and for assessing the performance of numerical weather prediction systems. Existing lidar-based MLH detection algorithms typically do not use the full resolution of the ceilometer, require manual feature engineering, and often do not enforce temporal consistency of the MLH. To address these limitations, a novel MLH detection approach has been developed based on deep learning techniques for image segmentation. The concept of our Deep-Pathfinder algorithm is to represent the 24-hour MLH profile as a mask and directly predict it from an image with lidar observations. Therefore, range-corrected signal data was obtained from Lufft CHM 15k ceilometers at five locations in the Netherlands that were part of the operational ceilometer network. Input samples of 224×224 pixels were extracted, each covering a 45-minute observation period. A customised U-Net architecture was developed with a nighttime indicator and MobileNetV2 encoder for fast inference times. The model was pre-trained on 19.4 million samples of unlabelled data and fine-tuned using 50 days of high-resolution annotations. Qualitative and quantitative results showed competitive performance compared to two benchmark models: the Lufft and STRATfinder algorithms. Existing path optimisation algorithms have good temporal consistency, but can only be evaluated after a full day of ceilometer data has been recorded. Deep-Pathfinder retains the advantages of temporal consistency but can also provide real-time estimates. This makes our approach valuable for operational settings, as real-time MLH detection better meets the requirements of users such as in aviation, weather forecasting and air quality monitoring.

1 Introduction

The atmospheric boundary layer is the lowest part of the troposphere which is influenced directly by meteorological mechanisms near the surface, including heat transfer, evaporation and transpiration, and terrain induced flow modification (Stull, 1988). One of the processes observed in the atmospheric boundary layer is the vertical mixing of air. The mixing layer height (MLH) marks the change from vertical mixing of air near the surface and the free atmosphere. MLH is not constant, but varies throughout the day and can range from less than 100 metres to a few kilometres, depending on a myriad of factors including for instance the climatological region, landscape, surface conditions and weather. Accurate estimates of the MLH are important for several applications and purposes. For example, a shallow mixing layer results in a larger concentration of air pollutants near the surface, affecting population health through increased risk of respiratory diseases. Similarly, MLH also affects the dispersion of greenhouse gases throughout the atmosphere for which accurate estimates are needed. Further, research on the parametrisation of the atmosperhic boundary layer remains essential to the further improvement of numerical weather prediction (NWP) systems (Edwards et al., 2020). Therefore, the availability of reliable MLH estimates would be useful to test and improve NWP accuracy.

The MLH is not easily and accurately identified in real-time. Existing methods for MLH detection are commonly based on (i) thermodynamic, (ii) wind and turbulence, or (iii) aerosol characteristics (Kotthaus et al., 2023). Thermodynamic methods, such as the parcel method (Holzworth, 1964) and the bulk-Richardson method (Vogelezang and Holtslag, 1996), use temperature and humidity profiles. These methods provide good baseline performance and are frequently used as a benchmark for the development of new methods. Methods based on wind or turbulence attempt to measure the height of the layer where buoyancy-driven or shear-driven turbulence takes place. The measurement of wind and turbulence can be performed using various instruments, including sodars, radar wind profilers and Doppler lidars. Finally, aerosol-based MLH methods attempt to observe the result of the mixing process via a proxy, as generally a rapid drop in aerosols can be observed beyond the top of the mixing layer. This phenomenon can be observed using ceilometers that employ the lidar (light detection and ranging) measurement principle, emitting short laser pulses and measuring the back-scattering by aerosols to eventually obtain estimates of particle concentrations at different altitudes. Since the incomplete optical overlap of many lidar systems results in a blind spot at low altitudes, estimating the height of shallow mixing layers can be challenging. Possibilities and limitations depend mainly on the instrument optical design, which will not be addressed here.

Not all methods for MLH detection take the temporal progression of the MLH into account. Point-based detection models (i.e., at a specific time) have the advantage that more labelled data is available for model fitting. However, these methods occasionally experience sudden jumps in the MLH profile from one layer to another. Several methods addressed this issue by reinforcing temporal consistency through path optimisation mechanisms from graph theory, initially developed by de Bruine et al. (2017), and subsequently further enhanced, such as in PathfinderTURB (Poltera et al., 2017), CABAM (Kotthaus and Grimmond, 2018) and STRATfinder (Kotthaus et al., 2020). For these type of approaches, it is common to reduce the temporal resolution of the input data to one or two-minute segments.

Some studies have developed approaches based on machine learning to further improve detection accuracy. For example, unsupervised methods such as cluster analysis have been used to detect the boundary layer based on backscatter data (Toledo et al., 2014). Further, Rieutord et al. (2021) compared the use of k-means clustering and AdaBoost. The accuracy of these two approaches varied substantially across measurement sites. However, the (initial) application of the machine learning methods showed potential and various suggestions for future research were made to further improve performance. Min et al. (2020) applied clustering algorithms for post-processing the results of several existing MLH detection algorithms. Extended Kalman filters were used by Lange et al. (2014) to model simplified statistics of MLH dynamics and the measurement noise. Further, Vivone et al. (2021) used edge detection techniques to identify layer boundaries. Allabakash et al. (2017) used fuzzy logic to combine the range-corrected signal-to-noise ratio, the vertical velocity, and the Doppler spectral width of the vertical velocity to identify MLH from a radar wind profiler. Bonin et al. (2018) also applied fuzzy logic to combine data from different scanning strategies of a Doppler lidar, determining where turbulent mixing is present. Various studies have also combined remote sensing information with other atmospheric variables. For example, gradient boosted regression trees were used by de Arruda Moreira et al. (2022) to predict the MLH estimated with microwave radiometer data based on the MLH estimated with ceilometer data and several atmospheric variables. Krishnamurthy et al. (2021) used the random forest algorithm to combine Doppler wind lidar MLH estimates using the method by Tucker et al. (2009) with various meteorological measurements such as surface relative humidity, air temperature, soil moisture, and turbulence kinetic energy. These approaches have been shown to generally improve prediction accuracy, although the use of multiple data sources may complicate the large-scale implementation in a real-time detection network.

In summary, several methodological challenges still remain. Few methods incorporate temporal information to avoid jumps between layers. For example, this is an issue for some of the machine learning methods described above. In contrast, many methods that are not based on machine learning require expert knowledge to manually set modelling thresholds (e.g., for nighttime detection, instrument- and site-specific tuning). This also extends to the manual specification of guiding restrictions for layer selection. An open research question for MLH detection is how to combine the advantages of different methods in a single approach. In particular, it would be beneficial to (i) promote temporal consistency of the MLH profile, (ii) use the full resolution of the ceilometer, and (iii) limit manual feature engineering, specification of rules for layer selection, and site-specific tuning parameters. Further, not all existing methods can be used for real-time detection, which is an important quality for operational use. These challenges form the basis for the Deep-Pathfinder MLH detection algorithm described in this paper.

2 Materials and methods

Our study proposes to process lidar data from ceilometers using computer vision techniques for image segmentation. Image segmentation has been used in many domains, including scene understanding for autonomous vehicles (Guo et al., 2021) and medical image analysis to detect various types of cancer (e.g., Dong et al., 2017). The concept of the new Deep-Pathfinder algorithm is to represent the 24-hour MLH profile as a mask (i.e., black indicating the mixing layer, white indicating the less turbulent atmosphere above) and directly predict the mask from an image with range-corrected ceilometer observations (see Fig. 1). This promotes temporal consistency of MLH estimates and limits manual feature engineering, while using the maximum resolution of the ceilometer.

Machine learning can estimate the link function between input and output images from historical data. Given the large-scale availability of ceilometer data and the high temporal and spatial resolution at which it is recorded, this domain is very suitable for machine learning approaches such as deep learning. However, the main challenge for a deep learning approach is the limited availability of annotated data. In particular, annotated data is generally not available for extended consecutive time periods, except for MLH estimates generated by other methods (e.g., ECMWF reanalysis). Further, annotated data is laborious to obtain, especially at high temporal resolution. Therefore, our research aimed to extract domain knowledge from unlabelled data, reducing requirements for the amount of annotated data.

https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

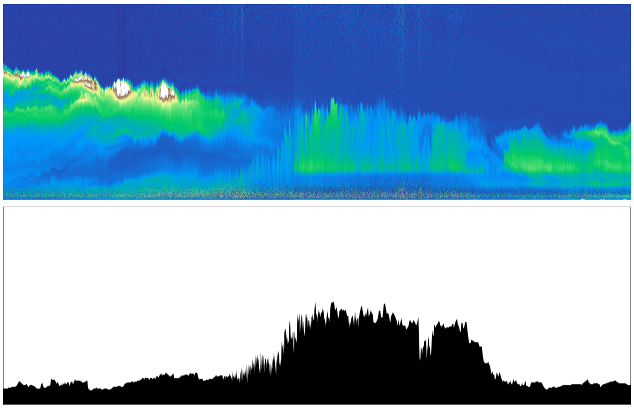

Figure 1. Concept of the Deep-Pathfinder model, showing an input image of the range-corrected lidar signal (top) and a corresponding mixing layer mask (bottom).

2.1 Data

A large dataset with unlabelled ceilometer data in NetCDF format (∼125 GB) was downloaded from the KNMI Data Platform (https://dataplatform.knmi.nl/). This dataset contained backscatter profiles from ceilometers in the KNMI observation network, recorded between June 2020 and February 2022. Data was available at five locations in the Netherlands: Cabauw, De Kooy, Groningen Airport Eelde, Maastricht Aachen Airport and Vlissingen. Throughout the observation period, each location operated a CHM 15k ceilometer from manufacturer Lufft, which is a one-wavelength backscatter lidar at 1064 nm. Data was recorded continuously, capturing the normalised range-corrected signal (RCS) at 12-second temporal and 10-meter vertical resolution. MLH estimates based on the manufacturer's algorithm were also available, but were only used as a benchmark.

2.2 Pre-processing and annotation

Various pre-processing steps were performed on the ceilometer data, using Python and OpenCV (Bradski, 2000). First, the RCS was capped to [0, 1e6] and rescaled to [0, 1]. The spatial range was cropped to a maximum altitude of 2240 meter, while retaining the 10-meter spatial resolution. A total of 7200 time steps was selected using the original temporal resolution of 12 seconds, capturing a 24-hour period of data. The resulting data was stored as a 16-bit grayscale image with 224×7200 pixels.

The software package labelme (Wada, 2022) was used for annotations (i.e., layer attribution). This tool enabled the creation of custom masks for image segmentation, including the export of selected points to JSON format. The resulting JSON data was converted to a black and white mask at the same 224×7200 pixels resolution as the input image. High-resolution annotations were created for 50 days in 2019, 2020 and 2021. The main location was Cabauw, as this location also provided humidity profile information to support nighttime annotations. Instead of creating a specialised model for the Cabauw ceilometer, the model should generalise to new locations. Therefore, several days of data from ceilometers at De Kooy and De Bilt were also

https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

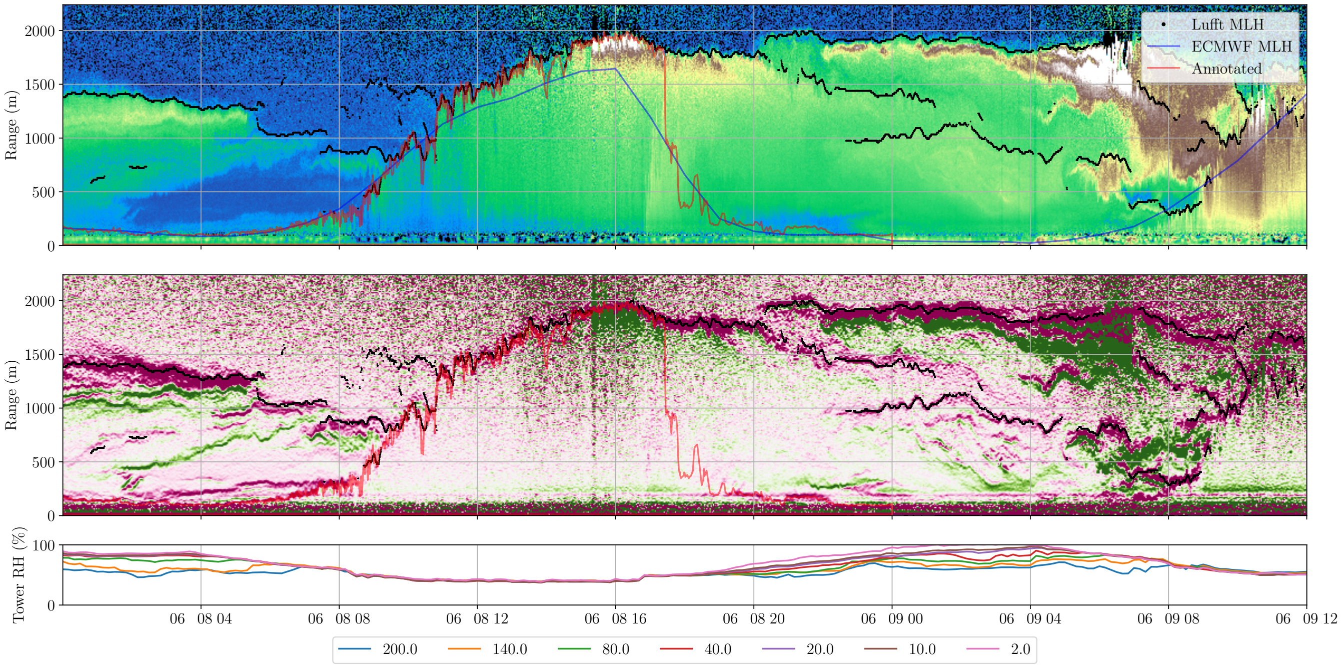

Figure 2. Combining data sources for annotation purposes, showing (i) layers detected by the ceilometer (black lines), (ii) the thermodynamic MLH from ECMWF (blue line), (iii) strong negative (magenta) and positive (green) gradients, and (iv) the measured relative humidity by the Cabauw mast at different altitudes (various colours).

annotated. Representative cases under a variety of atmospheric conditions were selected for annotation to cover a broad range of boundary layer dynamics.

The annotation process started with a visual inspection of the RCS data and corresponding gradient fields (see Fig. 2). Gradient estimation identified the location of layer boundaries in a consistent manner, leading to several candidate layers at some time steps. The information on potential layer boundaries was combined with other data sources to enhance the layer selection process. For example, thermodynamic MLH information from Cloudnet (ECMWF) model output (CLU, 2022) indicated the general atmospheric conditions. Further, MLH estimates from the manufacturer's layer detection algorithm were included for comparison purposes.

The layer attribution started with identification of the nocturnal MLH before sunrise by tracking backwards in time from the high-confidence MLH identified after sunrise. At this point, the humidity profile of the co-located 213-meter Cabauw mast (see Fig. 3) confirmed when the unstable boundary layer rose above 200 m (which generally coincided with the convective ejection patterns in the RCS). During nighttime the MLH may drop to very low altitudes into the partially visible range of the ceilometer, due to the use of biaxial optics in the CHM 15k ceilometer. Therefore, Cabauw mast measurements were used to aid in the identification of the nocturnal MLH. Humidity levels were similar when there was sufficient mixing, so low nocturnal MLHs can be approximated as the altitude where the humidity measurements started to diverge. For example, Fig. 2 indicates

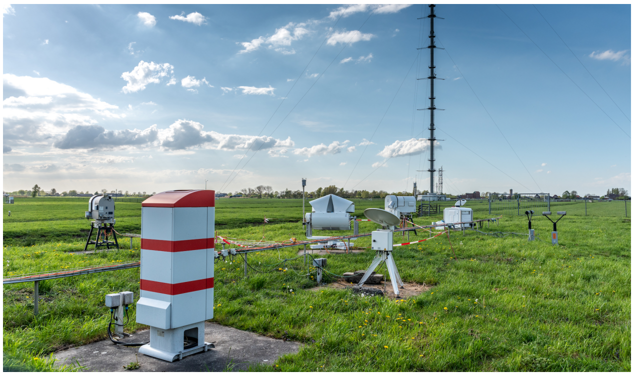

Figure 3. The ceilometer and 213-meter mast at the same location at Cabauw support high-quality annotations of the nocturnal MLH.

a MLH between 40 and 140 meters around 4am UTC on 8 June, which slowly rose to 200 meters around 7am. If the stable boundary layer of the following day was clearly visible in RCS data, it was followed backward in time to identify the formation of the nocturnal boundary layer after sunset.

Under less complex atmospheric conditions, the convective boundary layer plumes can often be identified in great detail from sharp layer edges. To take advantage of the high temporal and spatial resolution of the RCS, the MLH under such conditions was annotated following the ejection patterns inside the entrainment zone (EZ). Hence, the average MLH was located somewhere inside the EZ, while the amplitude between the local minimum and maximum MLH provided an indication of its thickness. The convective boundary layer was usually confidently annotated as the MLH until late afternoon or early evening, when the boundary layer was fully developed. For cases with convective clouds forming on the top of the boundary layer or low stratiform clouds with no clear aerosol layer underneath, the apparent cloud top height was annotated as the MLH. Periods with rain were annotated with a value of 0, as the MLH is undefined during precipitation.

The transition region from the unstable diurnal to the stable nocturnal MLH may not be clear from aerosol data (Wang et al., 2012). To complete the annotations in this region, the thermodynamic information informed the gradual decline towards the nocturnal MLH. Annotated layers were connected following a (low gradient) layer edge, where available. An example of this process can be seen in Fig. 3 between 4pm and 8pm. If the aerosol profile was too smooth, a sudden jump to the nocturnal boundary layer height was annotated.

Due to the physical processes leading to vertical mixing, the RCS profile shows distinct differences during the day and at night. To differentiate between the stable and convective boundary layer, a nighttime variable was included. Specifically, sunrise and sunset times in UTC were computed for the corresponding date and stored with the images using a nighttime variable. This assists the model in distinguishing whether an estimate of the stable or convective boundary layer is expected. In summary, one labelled sample consisted of a 24-hour pre-processed RCS image, a nighttime variable, and a corresponding annotated mask. All samples were converted to TensorFlow's TFRecord format for modelling purposes (Abadi et al., 2015).

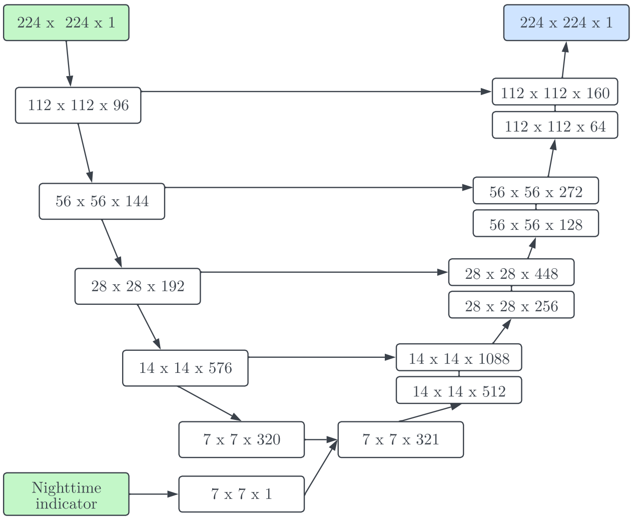

Figure 4. Deep-Pathfinder neural network architecture with inputs (green) and output (blue).

2.3 Model architecture

The Deep-Pathfinder algorithm is based on the U-Net architecture (Ronneberger et al., 2015), which is a frequently used model for image segmentation tasks. U-Net extracts features from an input image using consecutive convolutional layers (i.e., the encoder). From the latent space representation the dimensions are increased again to obtain an output mask (i.e., the decoder). Skip connections connect corresponding layers in the encoder and decoder to increase details in the generated mask. This generic U-Net architecture was adapted to the task of MLH detection. The encoder was based on MobileNetV2 (Sandler et al., 2018), which was originally developed for constrained compute environments such as mobile and embedded devices. This was chosen to ensure fast inference times for potential operationalisation, as MobileNetV2 was developed specifically for low latency inference. The input dimensions of the RCS image were 224×224 pixels from which a 7×7 block with 320 features was extracted after several subsequent layers. Further, the U-Net architecture was adapted to incorporate different boundary layer dynamics before and after sunset. Specifically, a nighttime indicator was added to the extracted features as an additional channel, indicating whether the sample mainly occurred inside or outside the sunrise to sunset window for the specific date and time of the sample. This resulted in a latent space with dimensions 7×7×321. The architecture used several transposed convolutional layers to decode the latent space and obtain a 224×224 pixels output image. For each pixel a single output value was produced, representing (i) the RCS value during pre-training, or (ii) a mixing layer indicator during transfer learning. A graphical representation of the neural network architecture is presented in Figure 4.

2.4 Model calibration

While annotated data is laborious to obtain, unlabelled data contains readily available and valuable information on the typical patterns observed in lidar signals. Therefore, part of the network was pre-trained to aid the calibration on limited annotated samples afterwards. Unsupervised pre-training was implemented by removing the skip connections and nighttime indicator from the neural network architecture to create an autoencoder with an equivalent structure. Removing the skip connections was necessary as information would otherwise flow directly from input to output without passing through the encoder/decoder structure. Unlabelled ceilometer data (see Sect. 2.1) was used to train this autoencoder, where the input RCS image was also used as the target image. In total, 19.4 million different samples of 224×224 pixels were extracted from the unlabelled data through cropping. Given the temporal resolution of 12 seconds, a total of 6976 different images can be extracted from a full day of data, each representing a period of almost 45 minutes. All samples used a fixed altitude range of 0 to 2240 meters, maintaining the original vertical resolution of 10 meters. Model calibration was performed with TensorFlow r2.6 using NVIDIA A100 GPUs of the Dutch National Supercomputer Snellius. The binary crossentropy loss function was used for model calibration in combination with the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 1e-3. After approximately 11 million iterations (about five days), the model reached convergence with a reconstruction loss close to 0. The pre-training task resulted in a calibrated encoder and decoder network. Only the encoder weights were retained to initialise the Deep-Pathfinder architecture with skip connections, extracting valuable features from RCS images without the use of annotated data. Unsupervised pre-training of encoder weights outperformed randomly initialising weights or loading ImageNet weights (results not presented).

Subsequently, transfer learning was used to fine-tune the pre-trained model for the task of mask prediction. During model calibration, each batch contained samples of one day randomly selected from the training data. This 24-hour RCS image and mask were randomly cropped to extract a batch of 16 corresponding image pairs of 224×224 pixels. Although optimal randomisation would be obtained if one batch contained samples from various days, this implementation choice ensured the GPU was fully utilised. For each 45-minute sample, it was assessed whether the majority of the sample concerned measurements between sunrise and sunset to set the value of the binary nighttime indicator. The deep learning model was provided with both the RCS image and the nighttime indicator as inputs, while the annotated mask was used as the target image. For illustration purposes, Fig. 5 shows several sample training image pairs. As 50 days of ceilometer data were annotated, the training set contained approximately 350,000 samples to fine-tune the deep learning model. Typically, an experiment required less than 50 epochs of training on labelled data for transfer learning. A small validation dataset for one additional annotated day was used to tune model hyperparameters (n = 1396 samples, including 765 daytime and 631 nighttime images). The validation set did not contain data from any of the days present in the training set. The main selection criterion for model evaluation was the mean accuracy of the generated masks in the validation set, providing a quantitative scoring mechanism for different experiments. In addition, predictions for a full validation day were visualised by creating a 24-hour prediction mask (see Sect. 2.5) to get additional insights on model behaviour. For example, models with comparable validation accuracy could show differences in smoothness of the decline in MLH estimates after sunset. This qualitative information provided secondary input in the model selection process, after candidate models were selected based on high validation accuracy. Finally, a test set was used with unused ceilometer data of the second half of 2020 at Cabauw, to obtain the unbiased performance of the final tuned model.

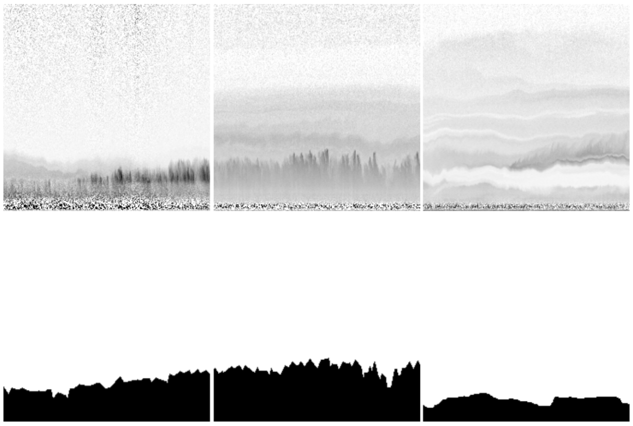

Figure 5. Three samples at Cabauw for (a) 15-Sep-2019 at 10:49am–11:34am, (b) 03-Sep-2021 at 11:33am–12:18pm, and (c) 14-Jun-2021 at 11:46pm–12:31am local time. Each sample consisted of a grayscale input image (top), a nighttime indicator (not shown) and annotated output mask (bottom).

2.5 Model inference and post-processing

The calibrated model generated masks corresponding to a 45-minute period, rather than providing a 24-hour sequence of MLH values. Therefore, several post-processing steps were required. First, output masks were generated via model inference using one-minute intervals throughout the entire day. This led to overlapping predictions for each time step, which were averaged to obtain a full 24-hour mask. Subsequently, an MLH profile was extracted from the 24-hour mask using the following method. Each predicted mask was processed column-wise, identifiying the MLH at time step t independent of other time steps. A loss function was formulated to evaluate the plausibility of every possible pixel $p \in \{1, 2, \dots, 224\}$ to represent the MLH, for a fixed time t. The loss penalised the number of pixels below p that were predicted as non-black, plus the pixels above p that were predicted as non-white. This was not based on the binary outcome for each pixel, but on the softmax model predictions (i.e., using a pixel-wise sigmoid function). MLH at time t was estimated as the value $\hat{p}$ that minimised this loss, multiplied by the spatial resolution of 10 metres.

Deep-Pathfinder performance was compared to MLH estimates from manufacturer Lufft (i.e., a proprietary algorithm of the ceilometer manufucturer based on wavelet covariance transform) and a state-of-the-art detection algorithm, STRATfinder (Kotthaus et al., 2020). At the time of publication, STRATfinder was still in active development and the STRATfinder data for this study was received from IPSL on 6 May 2022. Ceilometer data from the second half of 2020 at Cabauw was used as a test set, as during this period both Lufft and STRATfinder MLH estimates were available. This provided opportunities to place Deep-Pathfinder performance in perspective.

3 Results

3.1 Qualitative assessment

Fig. 6 shows the out-of-sample performance on new data not used for model calibration. Deep-Pathfinder estimates were compared to the benchmark methods for all days in the test set. The selected days in Fig. 6 contain varying conditions, including the typical growth of the convective boundary layer during the day, periods of precipitation, low clouds, hardly visible decay after sunset, multiple cloud layers, and a day without strong convection. The convective boundary layer during daytime was typically captured well by all three methods, with minimal differences in MLH between them. On 2 October and 6 August 2020 (Fig. 6, top row), the Lufft wavelet covariance transform algorithm jumps between several residual layers before sunrise. Deep-Pathfinder and STRATfinder correctly identified the nocturnal MLH around 100–200 meters altitude, although STRATfinder estimates were at a constant level slightly above the actual MLH due to guiding restrictions in the algorithm. Another difference between Deep-Pathfinder and STRATfinder was that Deep-Pathfinder followed short-term fluctuations in MLH more closely than STRATfinder due to the use of high-resolution input data. All algorithms had difficulties capturing the decline in MLH around sunset, which is a typical limitation for MLH detection based on aerosol observations. For example, for 6 August 2020 a sudden jump in MLH is visible for both Deep-Pathfinder and STRATfinder, although at a different time.

For complex atmospheric conditions, a considerable amount of MLH estimates of the Lufft algorithm were missing due to quality control flags. An example is provided for 9 July 2020, where Lufft estimates were only available after 8pm UTC and not during low cloud conditions. In most cases, Deep-Pathfinder and STRATfinder were still able to provide appropriate MLH estimates. In a few cases, STRATfinder predictions were missing due to quality control flags (e.g., 13 October 2020). During the precipitation event on 2 October 2020 around 7 to 9pm UTC, Deep-Pathfinder has been trained to predict 0 (i.e., not applicable), while Lufft predictions jumped to about 2,500 meter altitude. The example of 5 July 2020 shows that for multiple cloud layers Deep-Pathfinder and STRATfinder typically followed a different layer. When a clear convective boundary layer is not apparent (e.g., 10 December 2020), Deep-Pathfinder and STRATfinder were still able to correctly track the shallow MLH throughout the day.

3.2 Correlation analysis

A statistical assessment of overall agreement between the algorithms was performed through a correlation analysis. For each day in the July to December 2020 test period, the Pearson correlation was computed between the time series of each pair of algorithms. Deep-Pathfinder and STRATfinder scored an average correlation of 0.591 (see Table 1). In contrast, the Lufft algorithm obtained a substantially lower correlation with either method. Alignment of the algorithms was also not constant throughout the test period. Table 2 shows the distribution of the number of days the correlation was in a pre-defined range, indicating that on the majority of days the correlation between Deep-Pathfinder and STRATfinder was between $[0.6, 0.8)$ or $[0.8, 1]$.

https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

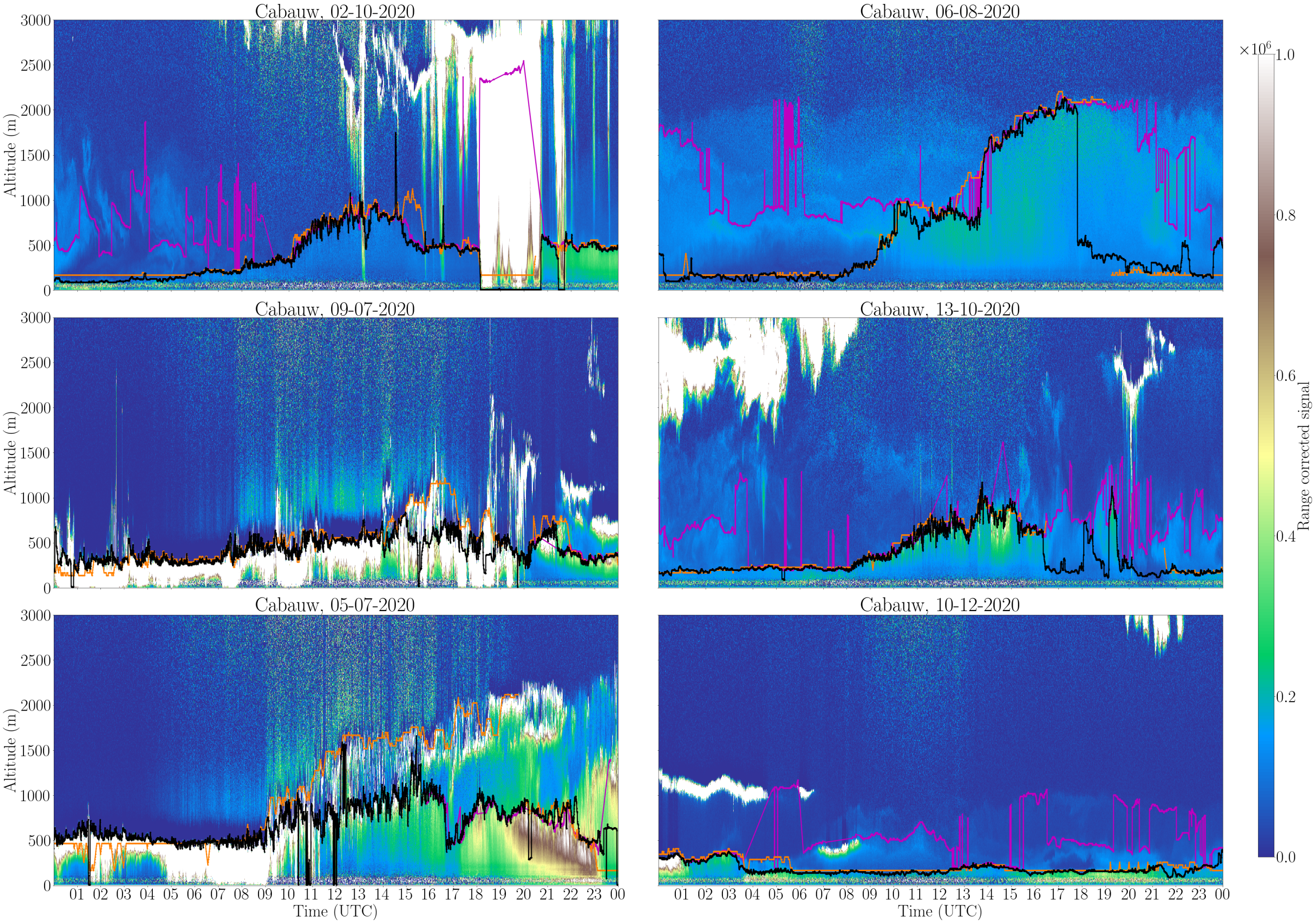

Figure 6. Performance comparison of Deep-Pathfinder (black), STRATfinder (orange) and Lufft (purple) on selected days at Cabauw.

Table 1. Mean Pearson correlation between daily time series of Deep-Pathfinder, STRATfinder and Lufft in the July to December 2020 test set.

	Deep-Pathfinder	STRATfinder	Lufft
Deep-Pathfinder	1	0.591	0.281
STRATfinder	0.591	1	0.184
Lufft	0.281	0.184	1

https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

Table 2. For each pair of algorithms, this table lists the total number of days that the Pearson correlation was within the specified ranges.

Correlation	Deep-Pathfinder vs. STRATfinder	Deep-Pathfinder vs. Lufft	STRATfinder vs. Lufft
$[-1.0,-0.8)$	0	0	0
$[-0.8,-0.6)$	0	4	5
$[-0.6,-0.4)$	0	2	7
$[-0.4,-0.2)$	1	10	19
$[-0.2,0.0)$	5	20	27
$[0.0,0.2)$	13	27	19
$[0.2,0.4)$	20	22	20
$[0.4,0.6)$	27	40	29
$[0.6,0.8)$	48	25	16
$[0.8,1.0]$	47	7	11

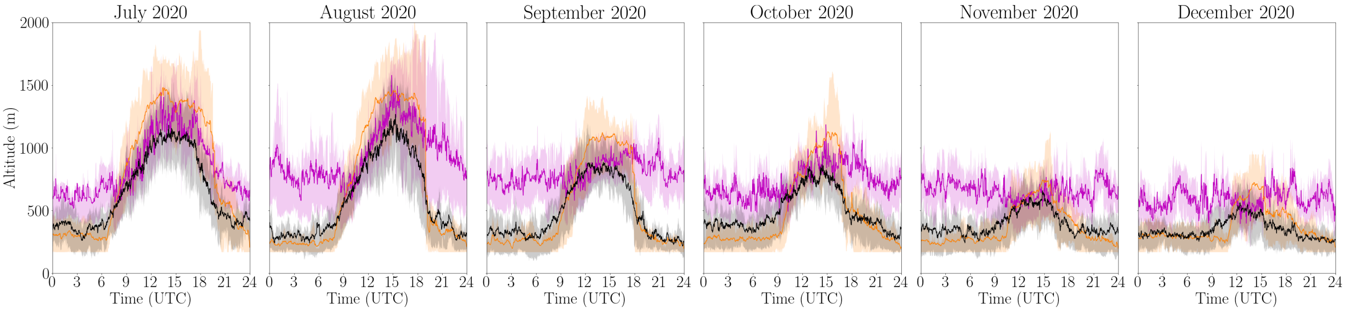

Figure 7. Diurnal MLH patterns per month at Cabauw for Deep-Pathfinder (black), STRATfinder (orange) and Lufft (purple), including interquartile ranges.

3.3 Diurnal MLH patterns

The performance of the MLH detection methods across different seasons can provide insights related to model robustness in terms of showing expected behaviours. Fig. 7 shows the mean MLH estimates throughout the day for the different algorithms, for each month in the test set. The interquartile range (i.e., 25th to 75th percentile) of the MLH estimates is also included in this figure. For consistency, the temporal resolution for this analysis was reduced to one minute for all methods. A gradual decline in peak MLH can be observed from July and August towards December. STRATfinder typically reached higher peak values than Deep-Pathfinder. In contrast, nocturnal MLH conditions were more stable throughout the months in the test set. There was no significant difference between STRATfinder and Deep-Pathfinder for the nocturnal MLH. However, the Lufft algorithm obtained higher nocturnal MLH estimates as it had a tendency to follow residual layers (see Sect. 3.1).

https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

The following two explanations were identified for the lower peak MLH of Deep-Pathfinder compared to STRATfinder. Most importantly, in case of multiple cloud layers our annotations typically followed the lower layer, while STRATfinder followed the higher layer. For example, this behaviour can be observed on 5-July-2020 (see Fig. 6). Secondly, Deep-Pathfinder MLH estimates fluctuated more by following short-term reductions in MLH, while STRATfinder did not. In Fig. 6, this behaviour can be observed between 10:00 and 13:00 on 13-Oct-2020. Both of these modelling choices led to a reduction in peak MLH for Deep-Pathfinder predictions, in comparison to STRATfinder.

3.4 Using alternative neural network architectures

The choice of neural network architecture is an important modelling choice in deep learning research. For example, various alternative neural network architectures have been developed based on U-Net. The latest architectures typically obtain higher performance on benchmark datasets than the standard U-Net implementation. Therefore, the performance of some of these new architectures has been investigated, quantifying the impact of the architectural choices on model performance and providing insights related to promising directions for future research. Specifically, the performance of Swin-Unet (Cao et al., 2021), UNet 3+ (Huang et al., 2020), Attention U-Net (Oktay et al., 2018), TransUNet (Chen et al., 2021), U^2-Net (Qin et al., 2020) and ResUNet-a (Diakogiannis et al., 2020) was investigated. Model implementations were obtained from the Keras UNet collection (Sha, 2021). Note that these experiments did not use any form of unsupervised pre-training. Therefore, the Deep-Pathfinder architecture without pre-training on unlabelled lidar data (referred to as ‘U-Net nighttime indicator’) was also included for comparison purposes. Further, a simpler architecture without nighttime indicator and no pre-training (‘U-Net base’) was included, as this indicator was not implemented for the alternative architectures.

After model training, masks were predicted for all samples in the validation set, and the MLH was extracted for a full day. Tables 3–5 provide statistical results on the mean absolute error (MAE), mean squared error (MSE) and Pearson correlation that were obtained. These statistics were computed with respect to the annotations for the validation set, which followed the same annotation process as described in Sect. 2.2. The results are indicative of the performance of different model architectures. Note that performance was only evaluated on the validation set, which was also used for model selection. A full evaluation on six months of test data was not performed for the alternative architectures.

Tables 3–5 show a large variation in statistics between different neural network architectures. The best performance was obtained by ResUNet-a and Deep-Pathfinder, followed by U^2-Net and TransUNet. Notably the ResUNet-a architecture obtained better results on the validation set than Deep-Pathfinder, although only for the MLH decay after sunset.

Mean absolute error was the lowest for the noctural MLH and convective MLH during daytime. MAE was substantially higher for the late afternoon decay in MLH, which was the main difficulty the models faced. Correlation was highest during daytime, as the forecasts with high temporal resolution followed the annotated convective boundary layer in the validation set very well.

The U-Net base architecture performed worse than most of the newer architectures, which was expected due to architectural improvements. Further, both the U-Net nighttime indicator and Deep-Pathfinder architectures performed substantially better than the U-Net base architecture. This shows the benefits of (i) incorporating sunrise and sunset information explicitly in the

Table 3. MAE for the MLH in metres, obtained using the validation set. Neural network architectures have been sorted based on overall score.

		Time of day (UTC)		
Architecture	Overall	0–8h	8–16h	16–24h
ResUNet-a	41.3	19.3	23.4	81.2
Deep-Pathfinder	64.2	19.0	23.3	150.4
U-Net nighttime indicator	74.5	29.4	27.9	166.2
TransUNet	82.6	31.1	75.5	141.2
U^2-Net	84.8	23.9	23.2	207.2
U-Net base	104.5	24.3	59.5	229.9
Attention U-Net	113.7	32.3	144.9	164.0
UNet 3+	125.1	28.4	166.1	180.9
Swin-Unet	242.7	31.8	401.6	294.7

Table 4. As Table 3, but for MSE.

		Time of day (UTC)		
Architecture	Overall	0–8h	8–16h	16–24h
ResUNet-a	7,683	1,163	1,530	20,362
Deep-Pathfinder	18,287	858	2,406	51,610
U-Net nighttime indicator	20,097	2,235	2,682	55,389
U^2-Net	21,313	2,046	1,727	60,183
TransUNet	25,154	4,731	24,337	46,403
Attention U-Net	38,882	2,960	54,620	59,076
UNet 3+	45,181	2,260	68,584	64,707
U-Net base	45,234	1,215	16,149	118,367
Swin-Unet	105,116	1,473	214,610	99,264

model and (ii) unsupervised pre-training on large-scale lidar data to improve feature extraction. For example, overall MAE decreased from 104.5 metres for the U-Net base model to 74.5 and 64.2 metres for the U-Net nighttime indicator and Deep-Pathfinder architectures, respectively.

Table 5. As Table 3, but for Pearson correlation.

		Time of day (UTC)		
Architecture	Overall	0–8h	8–16h	16–24h
ResUNet-a	0.96	0.80	0.99	0.85
Deep-Pathfinder	0.91	0.86	0.99	0.62
U^2-Net	0.90	0.70	0.99	0.63
U-Net nighttime indicator	0.90	0.65	0.99	0.57
TransUNet	0.89	0.49	0.88	0.71
Attention U-Net	0.82	0.58	0.76	0.51
UNet 3+	0.79	0.71	0.71	0.41
U-Net base	0.78	0.89	0.92	0.10
Swin-Unet	0.32	0.77	0.75	-0.40

4 Discussion

4.1 Annotations and model robustness

Labelling MLH is a complex and time-consuming task. Further, deep learning methods typically require a very large number of labelled samples. This combination of factors complicates the development of machine learning approaches for MLH detection. The issue of obtaining sufficient training samples was adressed in our study by unsupervised pre-training and extracting many 45-minute samples from a 24-hour mask through random cropping. Further, image segmentation architectures can be trained using relatively few annotated samples (Ronneberger et al., 2015), making it a suitable approach for this particular application.

Annotating MLH with high temporal resolution has several advantages. For example, the model will become more responsive to observed changes in MLH. Further, short-term fluctuations in MLH could be used to provide an estimate of the thickness of the entrainment zone (Cohn and Angevine, 2000), which cannot be provided by many algorithms. Combining measurements from different sensors, such as Doppler wind lidar, ceilometer and microwave radiometer could further improve the accuracy of annotations. Note that using only ceilometer data as model input allows for integration of the algorithm in existing Automatic Lidars and Ceilometers networks (e.g., E-Profile, see Haefele et al., 2016). However, including these additional sensor data sources as model input could also further increase the accuracy of MLH detection models (e.g., Kotthaus et al., 2023).

To explore model robustness, we have investigated training the model on other data sources than the manually annotated MLHs. Specifically, annotations for Payerne were obtained from MeteoSwiss (Poltera et al., 2017) to train the deep learning architecture (results not presented here). This experiment indicated it was possible to capture the important characteristics from alternative annotated datasets. The Deep-Pathfinder methodology was robust against differences in annotation methods, leading

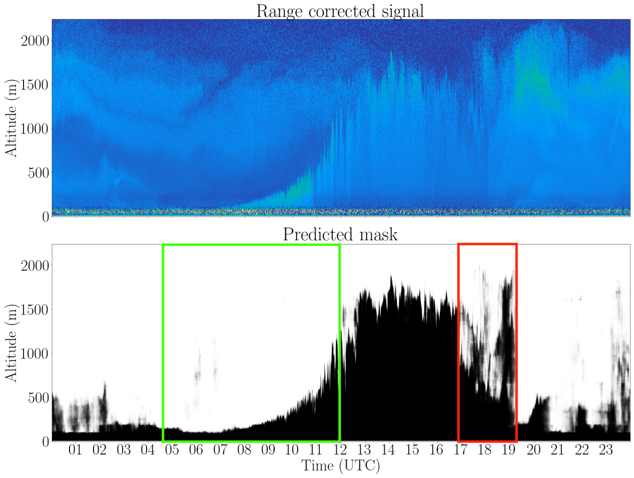

Figure 8. Model extensions: deriving quality control flags from unsharp regions in predicted masks. The green box provides an example of high model confidence, while the red box shows an example of lower model confidence.

to different results, but functioning appropriately regardless of the chosen dataset. Hence, the annotations and resolution of the input data mainly determine the quality of the final predictions.

For different types of ceilometers (e.g., Vaisala CL31), it is recommended to repeat the unsupervised pre-training using unlabelled data of the corresponding instrument. This should not be necessary when using Deep-Pathfinder at other locations with the same instrument type. Instead of unsupervised pre-training, the model could also be pre-trained using MLH predictions of (i) an existing MLH algorithm, (ii) a numerical weather prediction system, or (iii) synthetic data. This initial step would directly result in a base deep learning model for the task of mask prediction, which can be fine-tuned using limited high-quality labelled data. Fine-tuning can be an iterative process, where the current shortcomings of the model are used to slightly improve specific annotations in the training data. The model can then be retrained using the updated annotations to enhance performance.

4.2 Model extensions / future research

We have identified various model extensions that could be investigated in future research. For example, for model inference, each lidar observation appears in multiple time-shifted input samples. The overlap and averaging of prediction masks during post-processing leads to a grayscale output mask (see Fig. 8). Gray or blurry areas indicate model uncertainty and can be used to develop quality control flags for operational use. Specifically, the value of the loss during MLH extraction (see Sect. 2.5) could be considered as an indicator of model confidence. Further quality control flags could be set if rain is detected, since MLH is not clearly defined during periods of precipitation.

Instead of using only two output classes (i.e., mixing layer or not), image segmentation methods are suitable for the detection of multiple classes. Extending the Deep-Pathfinder algorithm to multi-class prediction would also be a valuable future

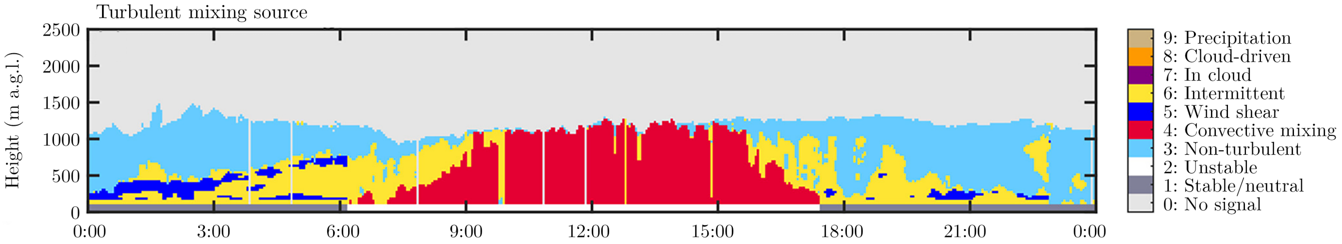

Figure 9. Model extensions: multi-class prediction (image from Manninen et al., 2018).

research direction. For example, Manninen et al. (2018) developed a method to obtain multiple classes from Doppler wind lidar information, such as convective mixing, non-turbulent, in cloud mixing, and wind shear (see Fig. 9). Obtaining such a set of annotated samples forms the main challenge for implementing this new functionality. In principle, besides some changes to the final layer, no major changes to the neural network architecture would be required.

The analyses with alternative neural network architectures indicate that there is potential to further improve model accuracy, especially since these models were not implemented with unsupervised pre-training or additional variables (i.e., nighttime indicator). U^2-Net has so many skip connections that it would not be feasible to apply the pre-training approach used in our study. In contrast, ResUNet-a performed well on the MLH detection task and only has a limited number of skip connections. A custom implementation of ResUNet-a with temporary removal of the six skip connections would allow for unsupervised pre-training. Hence, the ResUNet-a architecture is a promising candidate to further improve model accuracy, in future research. Note that accuracy was not the only consideration for choosing the deep learning architecture. ResUNet-a was 8.5 times slower to calibrate than Deep-Pathfinder because of the higher complexity. Computational efficiency is an important consideration for operational use.

5 Conclusions

Our study shows that computer vision methods for image segmentation can be adapted to successfully track layers in data recorded by ceilometers. Through the use of unsupervised pre-training on large-scale unlabelled lidar data, appropriate results for MLH estimation were obtained with only 50 days of annotations. Further, Deep-Pathfinder takes advantage of the full spatial and temporal resolution of the ceilometer, leading to high-resolution MLH estimates. One challenge for model development is that no ground truth MLH data is available, which also complicates method intercomparison. In comparison with existing MLH approaches (e.g., rule-based layer selection algorithms), the number of assumptions required for MLH detection was reduced. The initial structured annotation process (see Sect. 2.2) involves assumptions to determine the exact location of the MLH. However, manual feature engineering based on expert decisions is avoided, as the mapping between input and label is learned directly from large-scale data.

As shown in previous studies, layer attribution can be improved by taking into account temporal consistency. Although existing path optimisation algorithms have greatly improved the temporal consistency of MLH estimates, they can only be evaluated after a full day of ceilometer data has been recorded. Deep-Pathfinder retains the advantages of temporal consistency by assessing MLH evolution in 45-minute samples. However, our algorithm can also produce real-time estimates, by using the most recent 45 minutes of data and extracting the current MLH from the right-hand side of the output mask. The availability of real-time MLH estimates from a large-scale ceilometer network could be used for the advancement of NWP models via data assimilation. Finally, it makes a deep learning approach as presented here valuable for operationalisation, as real-time MLH detection better meets the requirements of operational users in aviation, weather forecasting and air quality monitoring.

Author contributions. All authors were involved in conceptualization of the study. DAG, JSW and AA jointly completed the data curation and annotation. JSW designed the methodology, performed the experiments, analysed results and wrote the original draft of the manuscript. AA, DAG and JWN reviewed and edited the manuscript. AA and JWN contributed to funding acquisition for this project.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This project was supported by the Dutch Research Council (NWO) through computation resources on the Snellius high-performance computing system (grant number EINF-3035) and by the Ruisdael Observatory for atmospheric science (grant number 184.034.015). The authors would like to thank Simone Kotthaus and Melania Van Hove (Institute Pierre Simon Laplace) for providing STRATfinder model predictions throughout 2020 at Cabauw, as well as Giovanni Martucci, Alexander Haefele (MeteoSwiss) and Yann Poltera (ETH Zurich) for providing annotated MLH estimates and corresponding ceilometer data at Payerne from the PathfinderTURB study.

https://doi.org/10.5194/amt-2023-80
Preprint. Discussion started: 2 May 2023

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., 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., and Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems, Software available from tensorflow.org, 2015.

Allabakash, S., Yasodha, P., Bianco, L., Venkatramana Reddy, S., Srinivasulu, P., and Lim, S.: Improved boundary layer height measurement using a fuzzy logic method: Diurnal and seasonal variabilities of the convective boundary layer over a tropical station, Journal of Geophysical Research: Atmospheres, 122, 9211–9232, https://doi.org/10.1002/2017JD027615, 2017.

Bonin, T. A., Carroll, B. J., Hardesty, R. M., Brewer, W. A., Hajny, K., Salmon, O. E., and Shepson, P. B.: Doppler lidar observations of the mixing height in Indianapolis using an automated composite fuzzy logic approach, Journal of Atmospheric and Oceanic Technology, 35, 473–490, https://doi.org/10.1175/JTECH-D-17-0159.1, 2018.

Bradski, G.: The OpenCV Library, Dr. Dobb's Journal of Software Tools, 2000.

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., and Wang, M.: Swin-Unet: Unet-like pure transformer for medical image segmentation, arXiv: 2105.05537v1 [eess.IV], https://doi.org/10.48550/ARXIV.2105.05537, 2021.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., and Zhou, Y.: TransUNet: Transformers make strong encoders for medical image segmentation, arXiv: 2102.04306v1 [cs.CV], https://doi.org/10.48550/ARXIV.2102.04306, 2021.

CLU: ECMWF, icon-iglo-12-23 model data, from Cabauw. Generated by the cloud profiling unit of the ACTRIS Data Centre, 2022, 2022.

Cohn, S. A. and Angevine, W. M.: Boundary layer height and entrainment zone thickness measured by lidars and wind-profiling radars, Journal of Applied Meteorology, 39, 1233–1247, https://doi.org/10.1175/1520-0450(2000)039<1233:BLHAEZ>2.0.CO;2, 2000.

de Arruda Moreira, G., Sánchez-Hernández, G., Guerrero-Rascado, J. L., Cazorla, A., and Alados-Arboledas, L.: Estimating the urban atmospheric boundary layer height from remote sensing applying machine learning techniques, Atmospheric Research, 266, 105 962, https://doi.org/10.1016/j.atmosres.2021.105962, 2022.

de Bruine, M., Apituley, A., Donovan, D. P., Klein Baltink, H., and de Haij, M. J.: Pathfinder: applying graph theory to consistent tracking of daytime mixed layer height with backscatter lidar, Atmospheric Measurement Techniques, 10, 1893–1909, https://doi.org/10.5194/amt-10-1893-2017, 2017.

Diakogiannis, F. I., Waldner, F., Caccetta, P., and Wu, C.: ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data, ISPRS Journal of Photogrammetry and Remote Sensing, 162, 94–114, https://doi.org/10.1016/j.isprsjprs.2020.01.013, 2020.

Dong, H., Yang, G., Liu, F., Mo, Y., and Guo, Y.: Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks, in: Medical Image Understanding and Analysis, edited by Valdés Hernández, M. and González-Castro, V., pp. 506–517, Springer International Publishing, Edinburgh, https://doi.org/10.1007/978-3-319-60964-5_44, 2017.

Edwards, J. M., Beljaars, A. C. M., Holtslag, A. A. M., and Lock, A. P.: Representation of boundary-layer processes in numerical weather prediction and climate models, Boundary-Layer Meteorology, 177, 511–539, https://doi.org/10.1007/s10546-020-00530-z, 2020.

Guo, Z., Huang, Y., Hu, X., Wei, H., and Zhao, B.: A survey on deep learning based approaches for scene understanding in autonomous driving, Electronics, 10, 471, https://doi.org/10.3390/electronics10040471, 2021.

Haefele, A., Hervo, M., Turp, M., Lampin, J.-L., Haeffelin, M., and Lehmann, V.: The E-PROFILE network for the operational measurement of wind and aerosol profiles over Europe, in: Proceedings of the WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (CIMO TECO 2016), pp. 1–9, World Meteorological Organization, Madrid, 2016.

Holzworth, G. C.: Estimates of mean maximum mixing depths in the contiguous United States, Monthly Weather Review, 92, 235–242, https://doi.org/10.1175/1520-0493(1964)092<0235:EOMMMD>2.3.CO;2, 1964.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., and Wu, J.: UNet 3+: A full-scale connected UNet for medical image segmentation, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1055–1059, IEEE, Barcelona, https://doi.org/10.1109/ICASSP40776.2020.9053405, 2020.

Kingma, D. P. and Ba, J. L.: Adam: a method for stochastic optimization, arXiv: 1412.6980 [cs.LG], https://doi.org/10.48550/arXiv.1412.6980, 2014.

Kotthaus, S. and Grimmond, C. S. B.: Atmospheric boundary-layer characteristics from ceilometer measurements. Part 1: A new method to track mixed layer height and classify clouds, Quarterly Journal of the Royal Meteorological Society, 144, 1525–1538, https://doi.org/10.1002/qj.3299, 2018.

Kotthaus, S., Haeffelin, M., Drouin, M.-A., Dupont, J.-C., Grimmond, S., Haefele, A., Hervo, M., Poltera, Y., and Wiegner, M.: Tailored algorithms for the detection of the atmospheric boundary layer height from common Automatic Lidars and Ceilometers (ALC), Remote Sensing, 12, 3259, https://doi.org/10.3390/rs12193259, 2020.

Kotthaus, S., Bravo-Aranda, J. A., Collaud Coen, M., Guerrero-Rascado, J. L., Costa, M. J., Cimini, D., O'Connor, E. J., Hervo, M., Alados-Arboledas, L., Jiménez-Portaz, M., Mona, L., Ruffieux, D., Illingworth, A., and Haeffelin, M.: Atmospheric boundary layer height from ground-based remote sensing: a review of capabilities and limitations, Atmospheric Measurement Techniques, 16, 433–479, https://doi.org/10.5194/amt-16-433-2023, 2023.

Krishnamurthy, R., Newsom, R. K., Berg, L. K., Xiao, H., Ma, P.-L., and Turner, D. D.: On the estimation of boundary layer heights: a machine learning approach, Atmospheric Measurement Techniques, 14, 4403–4424, https://doi.org/10.5194/amt-14-4403-2021, 2021.

Lange, D., Tiana-Alsina, J., Saeed, U., Tomás, S., and Rocadenbosch, F.: Atmospheric boundary layer height monitoring using a Kalman filter and backscatter lidar returns, IEEE Transactions on Geoscience and Remote Sensing, 52, 4717–4728, https://doi.org/10.1109/TGRS.2013.2284110, 2014.

Manninen, A. J., Marke, T., Tuononen, M., and O'Connor, E. J.: Atmospheric boundary layer classification with Doppler lidar, Journal of Geophysical Research: Atmospheres, 123, 8172–8189, https://doi.org/10.1029/2017JD028169, 2018.

Min, J.-S., Park, M.-S., Chae, J.-H., and Kang, M.: Integrated system for atmospheric boundary layer height estimation (ISABLE) using a ceilometer and microwave radiometer, Atmospheric Measurement Techniques, 13, 6965–6987, https://doi.org/10.5194/amt-13-6965-2020, 2020.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D.: Attention U-Net: learning where to look for the pancreas, arXiv: 1804.03999v3 [cs.CV], https://doi.org/10.48550/ARXIV.1804.03999, 2018.

Poltera, Y., Martucci, G., Collaud Coen, M., Hervo, M., Emmenegger, L., Henne, S., Brunner, D., and Haefele, A.: PathfinderTURB: an automatic boundary layer algorithm. Development, validation and application to study the impact on in situ measurements at the Jungfraujoch, Atmospheric Chemistry and Physics, 17, 10 051–10 070, https://doi.org/10.5194/acp-17-10051-2017, 2017.

Qin, X., Zhang, Z., Huang, C., Dehghan, M., Zaiane, O. R., and Jagersand, M.: U^2-Net: Going deeper with nested U-structure for salient object detection, Pattern Recognition, 106, 107 404, https://doi.org/10.1016/j.patcog.2020.107404, 2020.

Rieutord, T., Aubert, S., and Machado, T.: Deriving boundary layer height from aerosol lidar using machine learning: KABL and ADABL algorithms, Atmospheric Measurement Techniques, 14, 4335–4353, https://doi.org/10.5194/amt-14-4335-2021, 2021.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: convolutional networks for biomedical image segmentation, in: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, edited by Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., pp. 234–241, Springer International Publishing, Munich, https://doi.org/10.1007/978-3-319-24574-4_28, 2015.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C.: MobileNetV2: inverted residuals and linear bottlenecks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510–4520, IEEE, Salt Lake City, UT, https://doi.org/10.1109/CVPR.2018.00474, 2018.

Sha, Y.: Keras-unet-collection, https://github.com/yingkaisha/keras-unet-collection, https://doi.org/10.5281/zenodo.5449801, 2021.

Stull, R. B.: An introduction to boundary layer meteorology, vol. 13 of *Atmospheric and Oceanographic Sciences Library*, Springer, Dordrecht, 1 edn., https://doi.org/10.1007/978-94-009-3027-8, 1988.

Toledo, D., Córdoba-Jabonero, C., and Gil-Ojeda, M.: Cluster analysis: a new approach applied to lidar measurements for atmospheric boundary layer height estimation, Journal of Atmospheric and Oceanic Technology, 31, 422–436, https://doi.org/10.1175/JTECH-D-12-00253.1, 2014.

Tucker, S. C., Senff, C. J., Weickmann, A. M., Brewer, W. A., Banta, R. M., Sandberg, S. P., Law, D. C., and Hardesty, R. M.: Doppler lidar estimation of mixing height using turbulence, shear, and aerosol profiles, Journal of Atmospheric and Oceanic Technology, 26, 673–688, https://doi.org/10.1175/2008JTECHA1157.1, 2009.

Vivone, G., D'Amico, G., Summa, D., Lolli, S., Amodeo, A., Bortoli, D., and Pappalardo, G.: Atmospheric boundary layer height estimation from aerosol lidar: a new approach based on morphological image processing techniques, Atmospheric Chemistry and Physics, 21, 4249–4265, https://doi.org/10.5194/acp-21-4249-2021, 2021.

Vogelezang, D. H. P. and Holtslag, A. A. M.: Evaluation and model impacts of alternative boundary-layer height formulations, Boundary-Layer Meteorology, 81, 245–269, https://doi.org/10.1007/BF02430331, 1996.

Wada, K.: Labelme: image polygonal annotation with Python, https://doi.org/10.5281/zenodo.5711226, 2022.

Wang, Z., Cao, X., Zhang, L., Notholt, J., Zhou, B., Liu, R., and Zhang, B.: Lidar measurement of planetary boundary layer height and comparison with microwave profiling radiometer observation, Atmospheric Measurement Techniques, 5, 1965–1972, https://doi.org/10.5194/amt-5-1965-2012, 2012.