# Enhancing weather radar data by removing non-meteorological echoes, using neural networks trained on synthetic weather data

Richard Bölz[a,1,*], Tom Kirstein[a,1], Lukas Fuchs[a], Lukas Josipović[b], Annette Böhm[c], Ulrich Blahak[b], Volker Schmidt[a]

[a] *Institute of Stochastics, Ulm University, Ulm, Germany*
[b] *Deutscher Wetterdienst, Offenbach, Germany*
[c] *Deutscher Wetterdienst, Meteorologisches Observatorium Hohenpeißenberg, Hohenpeißenberg, Germany*

**Abstract**

Weather radars are essential for atmospheric research, weather forecasting and aviation safety, but they often detect non-meteorological echoes from scatterers such as insects, birds, and ground clutter. These non-meteorological echoes can then lead to misinterpretations in quantitative precipitation estimation and hydrometeor classification, which cause difficulties for atmospheric research and weather forecasting. This paper introduces a novel AI-based approach to identify such non-meteorological echoes in polarimetric radar data using a convolutional neural network. More specifically, we utilize a so-called U-net, which relies on large amounts of labeled radar data for training. To address the challenge of acquiring labeled radar data consisting of meteorological and non-meteorological echoes, we generate synthetic training samples by combining preprocessed winter data (meteorological echoes) with cluttered summer data (non-meteorological echoes) provided by Deutscher Wetterdienst. After training on synthetic data, evaluation of the U-net approach on operationally measured radar data shows that it outperforms a state-of-the-art method overall. This is particularly evident in the preservation of precipitation signals at the boundaries of larger weather events.

*Keywords and Phrases:* machine learning, convolutional neural network, synthetic data generation, meteorological radar data, polarimetric radar, non-meteorological echo detection

[1] RB and TK contributed equally.

[*] Corresponding author. E-mail address: `richard.boelz@uni-ulm.de`

# 1 Introduction

Since the late 1940s, weather radars have become indispensable tools in atmospheric and hydrological research and operational applications, e.g., research on clouds and precipitation formation processes, quantitative precipitation estimation (QPE), motion analysis and precipitation nowcasting, detection and tracking of convective storms, and radar data assimilation for numerical weather prediction (NWP), see Fabry (2015) for a broad overview. These weather radars are ground-based remote sensing instruments that scan an approximately conical volume originating from the radar site with high spatial and temporal resolution. Although this capability makes them highly suitable for atmospheric research, operational weather surveillance and forecasting, the data suffers from several sources of error. One particular problem is that radars not only detect meteorological particles, such as droplets, snowflakes, and hail, but also any other scatterers, i.e., objects in the scanned volume that scatter electromagnetic waves. This can lead to misinterpretation of the measurements during subsequent analyses. To address this problem, this paper proposes a novel AI-based method to identify signals corresponding to non-meteorological scatterers based on their spatial and statistical properties measured by weather radar products.

To explain the problem in more detail, it is useful to briefly summarize the basic measurement principle, as described in many textbooks; see, e.g., Doviak and Zrnić (1993); Sauvageot (1992). A short and highly directed pulse of electromagnetic waves is sent out via an antenna along a so-called ray. During a subsequent listening period, a time series of returned signals, also called echoes, from atmospheric scatterers is detected at the radar site. These echoes can be used to compute so-called radar moments, such as strength, phase, and frequency of the detected echoes, which contain information on the amount and nature of scatterers, as well as their radial movement towards or away from the radar site. The time elapsed since sending is a measure of the scatterers radial distance, or range, from the radar site. Thus, the sampling frequency of the time series defines the range resolution and the listening time defines the maximum measurement range. Each value of the time series represents the echoes from the scatterers contained in a certain measuring volume centered around a certain distance, the so-called range bin. For typical operational applications, the radar antenna is rotated at a constant angular velocity and records data along ray after ray in changing directions. The measuring volume of one range bin of one ray is referred to as a radar gate and the data from one full azimuthal rotation of the antenna is also referred to as a sweep. If the antenna's elevation angle is held constant for a sweep, it is called a plan position indicator (PPI) scan. Several successive PPI scans at different elevation angles are called volume scans, as they may sample large parts of the scanned volume, whose radius is given by the maximum range. Typical spatial resolutions of PPI scans are 250 m to 1 km in range and 1° in azimuth. Typical volume scans consist of 5 to 20 PPI scans at different elevations, all of which are measured within five minutes.

Modern dual-polarization radars can transmit and receive electromagnetic radiation at different polarization planes, most commonly horizontal (H) and vertical (V) (Bringi and Chandrasekar, 2001; Ryzhkov and Zrnić, 2019). Either the transmitted polarization plane is altered from pulse to pulse while both H and V channels are received simultaneously, or the H and V pulses are transmitted and received simultaneously, in the so-called simultaneous transmit and receive (STAR) mode. Most precipitation particles are asymmetric and on average they fall with their longer axis oriented horizontally. This preferential orientation is overlaid by more or less tumbling and wobbling, depending on the hydrometeor type. For example, snowflakes and pristine ice particles usually tumble and wobble more than oblate raindrops. This aerodynamic behavior causes differences in the returned signal strength and phase between the H and V pulses. These differences and their correlations in time, which depend on the hydrometeor shapes

and spatial orientations, led to the definition of other radar moments in addition to the signal strength, i.e., the radar reflectivity factors $Z_\mathrm{h}$ and $Z_\mathrm{v}$ at H and V polarization, respectively. For example, the differential reflectivity ZDR, given by $10\log_{10}(Z_\mathrm{h}/Z_\mathrm{v})$ (Seliga and Bringi, 1976), is related to the degree of particle asymmetry, and the differential phase shift $\phi_\mathrm{DP}$ is caused by differences in the propagation speed of the H and V waves, which depend on the shapes, the material, and the bulk density of the particles. Moreover, the linear depolarization ratio LDR pertains to asymmetric particles not being perfectly oriented horizontally or vertically, and the co-polar cross-correlation coefficient $\rho_\mathrm{HV}$ indicates the degree of variability in particle types and shapes within the measuring volume. Note that radars operating in STAR mode cannot measure LDR. These additional radar moments can be used to improve the echo interpretation with respect to the type and number of scatterers (Zhang, 2016), as will be further outlined in Sect. 2.1.

A problem for correct interpretation is the fact that the type of scatterers is not known a priori, and non-meteorological scatterers like insects or turbulent fluctuations of the air's refractive index (see Doviak and Zrnić (1993); Wilson et al. (1994)) might be misinterpreted as, e.g., precipitation. Besides non-meteorological echoes (which shall be the focus of the present paper) and attenuation, there are also many other systematic and statistical error sources to be taken into account, e.g., beam bending, beam blockage at obstacles, side lobe echoes, ground clutter, statistical signal fluctuations and radar miscalibration (Doviak and Zrnić, 1993; Fabry, 2015).

Biological scatterers, such as insects and birds, can provide valuable information to meteorological forecasters, and to ornithologists and entomologists. For example, so-called "fine lines" are sometimes visible in the radar data when thunderstorms expel cooler, denser air. Small airborne particles, such as dust and insects, are swept along the leading edge of this cooler air mass and, together with echoes caused by turbulent fluctuations of the air's refractive index (see below), contribute to the fine line (e.g., Wilson et al. (1994)), giving the forecaster valuable information about the characteristics of the thunderstorm. Radar data can also be used to track bird migrations (Hu et al., 2024) and monitor biodiversity (Bauer et al., 2024). However, for meteorological and hydrological applications such as QPE (Chen et al., 2021; Ryzhkov and Zrnić, 2019), precipitation nowcasting (Pulkkinen et al., 2020; Leinonen et al., 2023) or hydrometeor classification (Besic et al., 2016; Straka et al., 2000), it is essential to separate hydrometeor echoes from any other type of echo. This becomes more and more important because many processing steps are being automated, and human intervention based on individual experience is reduced.

Echoes from insects and turbulence in the boundary layer are predominantly observed during the summer. The turbulence echoes are commonly attributed to Bragg scattering, which occurs due to refractive index fluctuations at a length scale of half the radar wavelength (Doviak and Zrnić, 1993; Wilson et al., 1994). The air's refractive index depends on pressure, temperature and humidity, where mainly fluctuations in moisture and, to a lesser extent, temperature cause echoes. These echoes, together with those from insects, are centered around the radar's location because they are limited to lower altitudes within the planetary boundary layer (PBL) and the altitude of the radar ray increases with range. Hence, they appear as annular stationary structures of weak but non-negligible echoes. Radar moments prove to be very useful in distinguishing clutter from meteorological echoes because they provide additional information on the shape and uniformity of the scatterers. However, detecting and eliminating insect and turbulence echoes is still a challenge.

Conventional classification methods based on pre-defined thresholds of radar moments (Kilambi et al., 2018), Bayesian classifiers (Mäkinen et al., 2022), or fuzzy logic (Dufton and Collier,

2015; Radhakrishna et al., 2019) are difficult to configure such that they successfully filter out all echoes from insects in the boundary layer while preserving highly turbulent regions, the edges of true precipitation regions and weak precipitation signals. Especially challenging situations include (i) cases with snow echoes or regions in the melting layer, because they can exhibit polarimetric signatures similar to those of insects, (ii) regions far away from the radar sites, where the signal-to-noise ratio is low, and (iii) "mixed" situations, e.g., summertime convection surrounded by cloud-free regions with insects in the boundary layer. The classifiers mentioned above also depend on manually chosen input features. Finding features that capture the relevant spatial patterns, such as the annular structure of the insects, is difficult.

Artificial intelligence (AI) methods have been used in recent studies to detect and classify echoes from biological scatterers in radar data for the purposes of biological studies. The focus of these studies was on obtaining information about insects or birds and their subclasses, rather than retaining all meteorologically useful data. For example, in Jatau et al. (2021), a ridge classifier was applied to polarimetric radar data to distinguish between birds and insects, with meteorological signals filtered out before classification. Moreover, random forest approaches based on polarimetric radar data were used to discriminate between different biological scatterers (Gauthreaux and Diehl, 2020; Hu et al., 2024). Besides this, in Atanbori et al. (2025), the problem of noisy or uncertain training labels was addressed by iteratively replacing the labels of the initial training data with the high-confidence predictions of an ensemble of neural networks.

Unlike the above-mentioned conventional approaches that rely on predefined thresholds, Bayesian classifiers, or fuzzy logic, AI-methods can learn intricate patterns directly from data without explicit feature engineering. This adaptability makes them particularly promising for handling the complex and high-dimensional nature of radar data. Convolutional neural networks (CNNs) specifically excel at extracting spatial features from image-like data structures, such as radar sweeps, enabling tasks like object detection and segmentation, see Goodfellow et al. (2016). Among CNNs, the U-net introduced in Ronneberger et al. (2015) is particularly well-suited for segmentation tasks due to its encoder-decoder architecture that captures both local features and global context via skip connections. For further applications of CNNs in radar-based traffic and biomedical monitoring, see Kern et al. (2022); Krauss et al. (2024). In this context, the sweeps can be viewed as multi-channel images, where each pixel corresponds to a radar gate and each image channel corresponds to a radar moment.

The goal of this paper is to develop and test a new method to identify non-meteorological echoes in sweeps using a U-net and suitable radar moments as input. However, applying U-nets generally requires a large and diverse dataset of accurately labeled training samples. Labeling radar data is especially challenging, as the labeling of even an individual pixel requires expert consideration of multiple radar moments and their spatial arrangement in multiple surrounding pixels. Given that each training sample consists of thousands of pixels and training requires hundreds or even thousands of labeled training samples, relying on manually labeled training data is laborious. In the present paper, although a small number of manually labeled sweeps was utilized for testing and benchmarking, training was instead conducted on synthetically generated training samples. More specifically, we first identified sweeps that feature either no meteorological echoes or no non-meteorological echoes. Sweeps that have no meteorological echoes were identified by considering time periods when no precipitation was measured. To identify sweeps featuring a minimum amount of non-meteorological echoes, a time period was manually selected through expert evaluation. These initial sweeps were then combined to generate "mixed" synthetic training samples featuring both meteorological and non-meteorological echoes with known ground truth.

The remainder of this paper is organized as follows. Section 2 describes the acquisition of radar data and its preprocessing, as well as the architecture and training of the neural network that is used for the detection of non-meteorological echoes. Section 4 describes the results obtained by the presented method applied to a test dataset not used during training. On this test dataset, the network's results are compared to those obtained by the state-of-the-art (SotA) method at Deutscher Wetterdienst (DWD) described in Sect. 3. The results are discussed in Sect. 5, and Sect. 6 presents the conclusions.

# 2 Materials and methods

We now explain how the polarimetric radar data is acquired and represented, and introduce the polarimetric radar moments in more detail. Section 2.2 describes the preprocessing steps applied to the data. The preprocessed data is then used to generate the input for the neural network, the architecture of which is described in Sect. 2.3. Section 2.4 explains the synthetic training data generation and the network training is described in Sect. 2.5. Section 2.6 describes the model selection and early stopping, and Sect. 2.7 introduces the metrics used for the quantitative evaluation.

## 2.1 Radar data

The radar data used in this paper is provided by DWD, which operates a radar network of 17 dual-polarization weather radars, covering the entire landmass of Germany. These radars operate in the STAR mode as described above, where H and V polarized pulses are transmitted simultaneously and where the power and phase of the received signal are recorded simultaneously for H and V polarizations. The recorded signals are acquired in sweeps, each containing the data from one 360° revolution of the antenna, with the antenna first pointing north and then rotating clockwise. More precisely, every five minutes, each of the radars performs a volume scan consisting of 10 sweeps acquired at different elevation angles, as well as a terrain-following sweep (Helmert et al., 2014), which measures precipitation as closely to the ground as possible. In the following, we utilize only these terrain-following sweeps. If the radar view is unobstructed, the elevation angle at which the electromagnetic waves are sent out for such sweeps is set to 0.8°. In cases where orography or obstacles obstruct the view, the elevation is raised accordingly.

The radars scan at a range resolution of 250 m and an azimuthal resolution of 1°, with a maximum range of 150 km, resulting in 360 rays of 600 radar gates each. Each of these radar gates is sampled by 50 pulses of electromagnetic waves, which serve as input to compute various radar moments for this radar gate. An important step after collecting the radar measurements is quality control of the radar data, which provides a solid data base for subsequent quantitative precipitation estimation, hydrometeor classification and thunderstorm detection. One main goal of this process is to identify and subsequently filter out, or correct for, non-meteorological echoes, that is, the signals returned by non-meteorological scatterers. Typical non-meteorological scatterers include ground clutter (trees, mountains), biological scatterers (birds, insects), wind turbines, towers, ships, airplanes, external signals (e.g. wifi), and chaff. In contrast, meteorological scatterers are any hydrometeors, e.g., rain, snow, hail, or drizzle. The signals returned by meteorological scatterers are called meteorological echoes.

The various radar moments are particularly useful for this purpose of identifying and filtering out non-meteorological echoes, as they provide insight into different qualities of the detected scatterers. The most widely used radar moment is the radar reflectivity factor (Straka et al.,

2000; Gauthreaux and Diehl, 2020), which quantifies the strength of the returned signal. It is measured separately at horizontal and vertical polarization, yielding $Z_{\mathrm{h}}$ and $Z_{\mathrm{v}}$, respectively. Due to variations in hydrometeor type (liquid, frozen), size and number concentration, the radar reflectivity factor can vary by several orders of magnitude. Hence, it is usually expressed in logarithmic units (dBZ) and denoted by DBZH and DBZV, respectively, where the minimum detectable value increases with range. For example, at a distance of 1 km from the radar, the minimum detectable reflectivity factor is around $-38$ dBZ, whereas at ranges beyond 25 km, signals below $-10$ dBZ cannot be detected by DWD radars (Frech et al., 2017), i.e., they cannot be distinguished from noise.

The differential reflectivity ZDR contains information about the shape of the scatterers and is commonly used, e.g., to discriminate between liquid and frozen hydrometeors. In combination with other radar moments, it is also useful for identifying clutter.

Finally, because the linear depolarization ratio LDR cannot be measured directly by DWD radars, an estimated depolarization ratio (denoted by UDR) is deduced from the differential reflectivity ZDR and the co-polar cross-correlation coefficient $\rho_{\mathrm{HV}}$ (Melnikov and Matrosov, 2013). This radar moment is a measure of how homogeneous and spherical the scatterers are within a radar gate, and thus serves as a good discriminator between clutter and meteorological echoes (Kilambi et al., 2018; Michelson et al., 2020), with small values ($\ll 0$) indicating weather and values closer to 0 indicating clutter or biological signals. In the available DWD data, the reflectivity factors DBZH and DBZV, and the differential reflectivity ZDR have been filtered by a Doppler filter (Doviak and Zrnić, 1993) aiming to remove any stationary non-meteorological scatterers, whereas no such filter has been applied to the depolarization ratio UDR.

The DWD provided all terrain-following sweeps measured at the 17 different dual-polarization weather radars for three one-hour time periods: 09:00–10:00 UTC on June 16, 2022, 10:00–11:00 UTC on June 18, 2022, and 17:00–18:00 UTC on January 9, 2025. Radar sweeps collected in January 2025 are referred to as winter sweeps, and radar sweeps collected in June 2022 are referred to as summer sweeps. Moreover, the radar sweeps collected on June 16 feature both meteorological and non-meteorological echoes and are referred to as mixed summer sweeps. In contrast, the radar sweeps collected on June 18 were collected during a time period when no precipitation was measured, indicating that only non-meteorological echoes were detected. Therefore, these sweeps are referred to as cluttered summer sweeps. For the winter sweeps, the specific time frame was chosen with the goal of selecting sweeps featuring a minimum amount of non-meteorological echoes. However, even for the winter sweeps, a certain amount of non-meteorological echoes cannot be avoided. Figure 1 shows examples of the radar moments DBZH, ZDR and UDR obtained from different sweeps measured during these time periods.

The winter sweeps contain a variety of precipitation features (Fig. 2a), with isolated showers over north-western Germany, widespread stratiform precipitation in north-eastern Germany, a convective line over southern Germany, and some very light precipitation similar to insect signatures. The cluttered summer sweeps show different insect signatures for different radars (Fig. 2b): the radars in the south and east of Germany show strong insect signatures, whereas the radars in the north and north-west show lighter signatures.

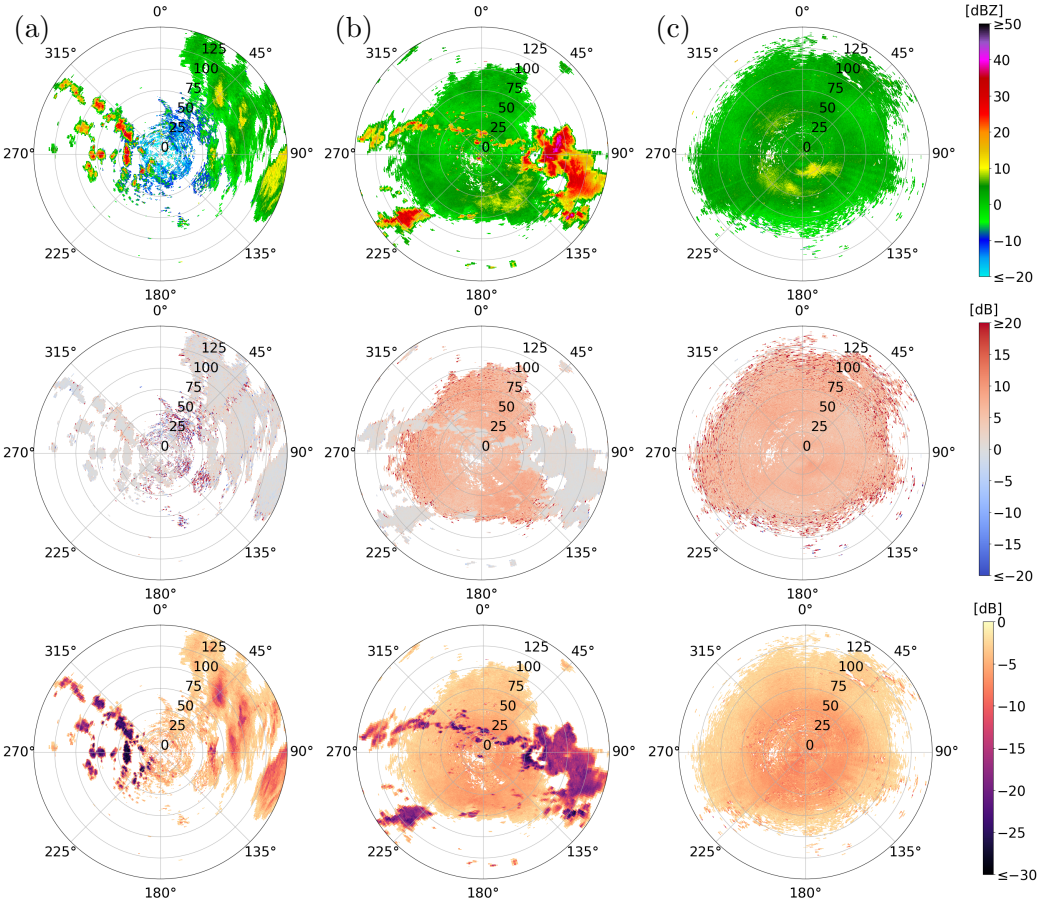


Figure 1: Horizontal reflectivity (first row), differential reflectivity (second row), and depolarization ratio (third row) for examples of (a) a winter sweep, (b) a mixed summer sweep, and (c) a cluttered summer sweep. White pixels are filtered by the initial preprocessing done by DWD. The sweeps shown were measured (a) on January 9, 2025, at 17:55 UTC in Flechtdorf, (b) on June 16, 2022, at 09:00 UTC in Dresden and (c) on June 18, 2022, at 10:00 UTC in Isen.

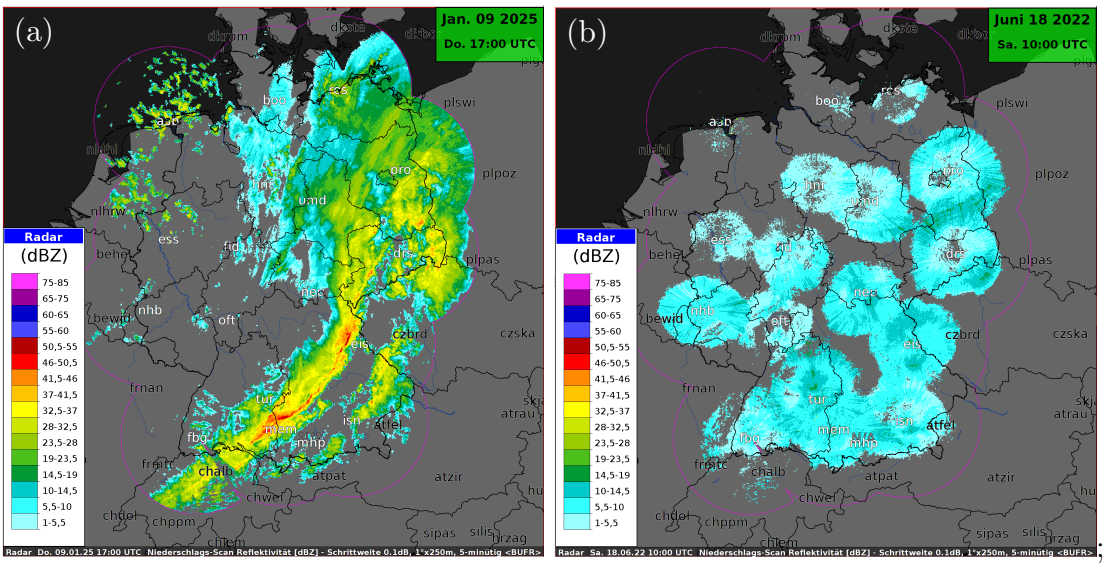


Figure 2: Composite radar images of the horizontal reflectivity over Germany: (a) the precipitation during the winter period, on January 9, 2025, at 17:00 UTC, and (b) the non-meteorological echoes during the cluttered summer period, on June 18, 2022, at 10:00 UTC.

## 2.2 Preprocessing of radar data

For each radar station, a single sweep consists of the measured values of the radar moments at the radar gates covered by the radar. These radar gates partition the scanned area into a polar grid defined by the radar's number of rays $n_\mathrm{a} \in \mathbb{N} = \{1, 2, \ldots\}$ and the number of range bins $n_\mathrm{r} \in \mathbb{N}$. More specifically, the radar gate corresponding to the $i$-th range bin of the $j$-th ray is defined as the set

$$g_{i,j} = \left\{ wr \begin{pmatrix} \sin\alpha \\ \cos\alpha \end{pmatrix} : \quad r \in \big(i-1, i\big], \alpha \in \left(\frac{2\pi}{n_\mathrm{a}}(j-1), \frac{2\pi}{n_\mathrm{a}} j\right] \right\} \subset \mathbb{R}^2, \tag{1}$$

where $w = 250$ denotes the range resolution of the radar in meters, $r$ is the distance from the radar site in units of the range resolution $w$, and the azimuthal angle $\alpha$ is measured clockwise from north.

Let $\mathcal{M}$ denote the set of radar moments available for a given sweep. Then, for each radar moment $m \in \mathcal{M}$, its measured values from the sweep can be represented by a single-channel image $A^m = (A^m_{i,j}) \in \mathbb{R}^{n_\mathrm{r} \times n_\mathrm{a}}$. For each pixel $(i, j) \in \Omega = \{1, \ldots, n_\mathrm{r}\} \times \{1, \ldots, n_\mathrm{a}\}$, the entry $A^m_{i,j}$ corresponds to the value of the radar moment $m$ at the associated radar gate $g_{i,j}$. Single-channel images can be combined into a multi-channel image $A = (A^m)_{m \in \mathcal{M}'} \in \mathbb{R}^{C \times n_\mathrm{r} \times n_\mathrm{a}}$, where $\mathcal{M}' \subseteq \mathcal{M}$ is the subset of included radar moments and $C = |\mathcal{M}'|$ is the number of channels. In the following, we refer to these multi-channel images as radar images. Moreover, the provided data also contains information on which radar gates had already been identified as non-meteorological during the preprocessing performed by DWD. For these radar gates, the corresponding pixels of the DBZH image are assigned a placeholder value $p = -40\,\mathrm{dBZ}$, which lies far below the reflectivity range of meteorological echoes. Pixels with a horizontal reflectivity value of $-40\,\mathrm{dBZ}$ are referred to as excluded pixels, as they conclusively do not correspond to meteorological echoes. All other pixels are referred to as included pixels. A Cartesian representation of a DBZH image is shown in Fig. 3a. Most of the radar images still contain pixels that likely correspond to non-meteorological echoes, even for radar images measured in winter. To filter some of these pixels out, we apply further preprocessing steps.

For example, as depicted in Fig. 3a, the DBZH images show weak signals with horizontal reflectivity values below $-10\,\mathrm{dBZ}$ near the radar. However, only very light drizzle could produce reflectivity values of that magnitude. That is, thresholding at $-10\,\mathrm{dBZ}$ most likely removes noise and clutter, and does not remove any convectively significant weather. Therefore, pixels with lower horizontal reflectivity values are considered to correspond to non-meteorological echoes and are thus excluded as potential meteorological echoes (Fig. 3b). Their DBZH values are then also set to $-40\,\mathrm{dBZ}$ to mark them as excluded pixels.

Moreover, meteorological echoes observed in the radar images typically form connected components consisting of several pixels (Fig. 1). In contrast, isolated included pixels are assumed to correspond to non-meteorological echoes. For this reason, each DBZH image is further processed by a despeckle filter consisting of two steps. First, we use the Hoshen-Kopelman algorithm (Hoshen and Kopelman, 1976) to identify all connected components of included pixels. More specifically, we apply the algorithm to the binarized image obtained by thresholding at $-40\,\mathrm{dBZ}$, separating the excluded pixels from the included ones. Here, we use the 4-neighborhood, which means that two pixels $(i_1, j_1), (i_2, j_2) \in \Omega$ with $(i_1, j_1) \neq (i_2, j_2)$ are neighboring if their associated radar gates $g_{i_1,j_1}, g_{i_2,j_2}$ share a common edge. Note that the DBZH image is periodic in the second component, see Eq. (1), since each index $j \in \{1, \ldots, n_\mathrm{a}\}$ represents the azimuthal angle of a ray, and the image corresponds to a sweep. Consequently, pixels corresponding to the first and last ray are neighboring if they have the same range index $i \in \{1, \ldots, n_\mathrm{r}\}$.

More precisely, for each pixel $(i, j) \in \Omega$, the set $N_{(i,j)} \subset \Omega$ of neighboring pixels can be written as $N_{(i,j)} = \{(k, \ell) \in \Omega : |i - k| + d(j, \ell) = 1\}$, where $d\colon \{1, \ldots, n_\mathrm{a}\} \times \{1, \ldots, n_\mathrm{a}\} \to \mathbb{N}_0$ denotes the periodic distance measure, given by

$$d(j, \ell) = \min\{|j - \ell|, n_\mathrm{a} - |j - \ell|\}. \tag{2}$$

for each $\ell \in \{1, \ldots, n_\mathrm{a}\}$. After identifying the connected components, those consisting of less than six undetermined pixels are selected. Then, the DBZH values of pixels $(i, j) \in \Omega$ in these components are set to $-40\,\mathrm{dBZ}$, indicating that they do not correspond to meteorological echoes. The result of the despeckling process is shown in Fig. 3c.

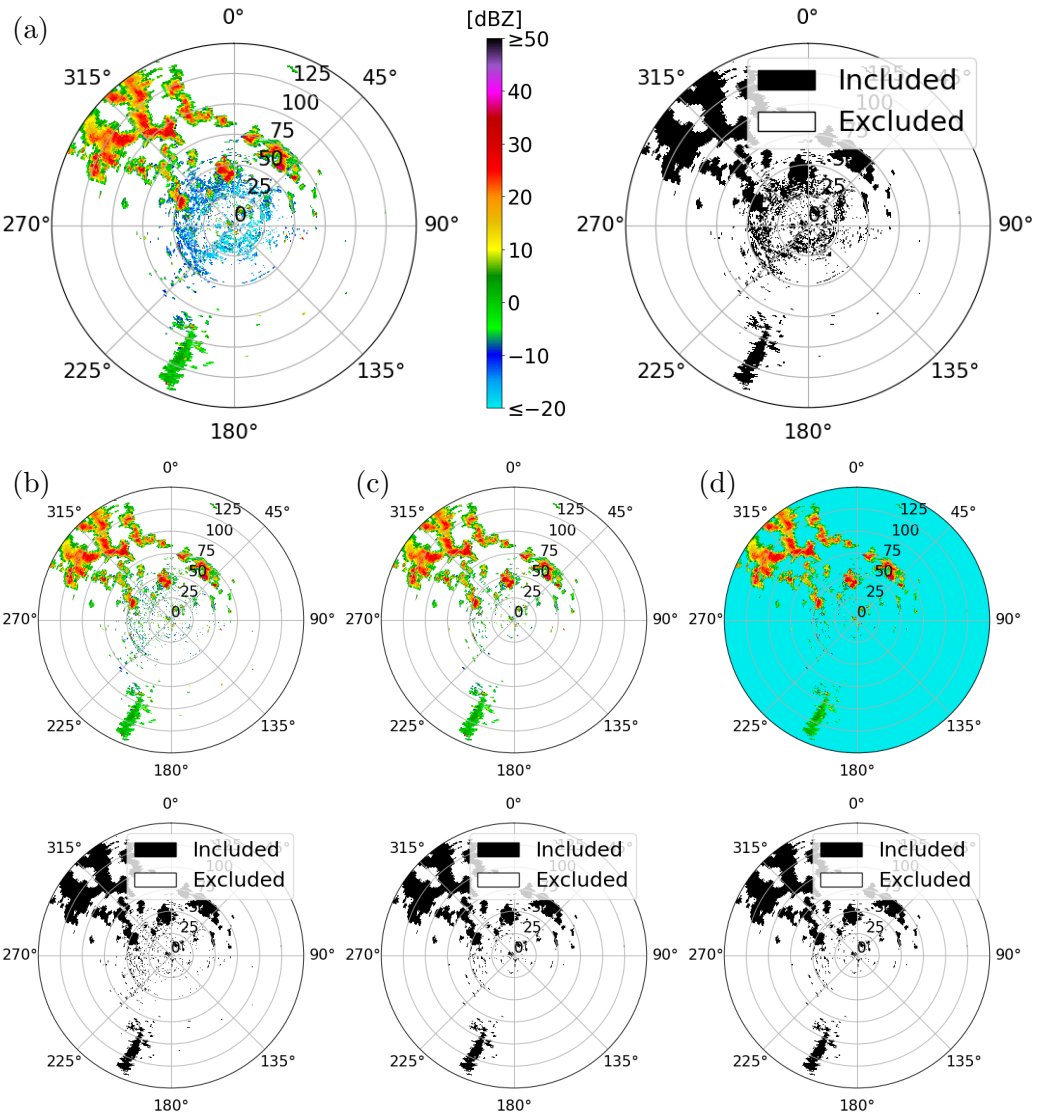


Figure 3: Cartesian representations of DBZH images: (a) DBZH image measured in winter, as provided by DWD, (b) after thresholding, (c) after despeckling, and (d) final DBZH image with placeholder values. White pixels in the DBZH images are either filtered out during the preprocessing by DWD or are assigned a placeholder value. The radar image shown was measured on January 9, 2025, at 17:00 UTC in Essen.

For the excluded pixels $(i, j) \in \Omega$, their corresponding entries in the DBZH image were set to $-40\,\mathrm{dBZ}$ to indicate that they conclusively do not correspond to meteorological echoes (Fig. 3d). For the other radar moments, ZDR and UDR, we also assign specific placeholder values

$p \in \mathbb{R}$ to the entries of their single-channel images. Similar to the horizontal reflectivity, these placeholder values are chosen to lie outside the range of typical values for meteorological echoes in our dataset. More specifically, for the differential reflectivity, the placeholder value is set to $p = 20\,\mathrm{dB}$. This particular value was chosen since it lies above typical values for meteorological echoes (around $0\,\mathrm{dB}$), but within the range for non-meteorological echoes (between 5 and $20\,\mathrm{dB}$) in our data, as can be seen in Fig. 1b and c. Furthermore, as described in Kilambi et al. (2018), except for some special cases such as hail, the depolarization ratio of meteorological scatterers is typically below $-12\,\mathrm{dB}$. In contrast, the depolarization ratio of non-meteorological scatterers typically ranges between $-12$ and $0\,\mathrm{dB}$. Therefore, we chose a placeholder value of $p = 0\,\mathrm{dB}$ for the depolarization ratio.

The presented preprocessing steps remove most of the clutter with low reflectivity values. However, many mixed summer radar images contain much larger connected components corresponding to more reflective non-meteorological scatterers, as can be seen in Fig. 1b. Detecting such echoes presents a challenging task for conventional methods such as fuzzy logic filtering (Dufton and Collier, 2015) or depolarization ratio thresholding (Kilambi et al., 2018).

## 2.3 Description of the network architecture

The task of detecting non-meteorological echoes can be viewed as a segmentation task of multi-channel two-dimensional (2D) radar image data. Each pixel of the image is classified as corresponding to meteorological or non-meteorological echoes, that is, each pixel is assigned a binary class label $c \in \{0, 1\}$. In Ronneberger et al. (2015), the so-called U-net architecture, a fully convolutional neural network, has been introduced to address similar segmentation tasks of 2D images.

A U-net has an encoder-decoder architecture with skip connections, consisting of a so-called contracting path that downsamples the spatial resolution of the features and an expansive path that upsamples it again. Both the downsampling and the upsampling paths consist of several levels. The network considered in the present paper is an adapted version of the U-net architecture comprising five levels, with the fifth level acting as a bottleneck. Consequently, the network performs four downsampling steps and four upsampling steps. Each downsampling step consists of two convolution blocks and a $2 \times 2$ max-pooling layer for downsampling the spatial resolution of the feature map (Goodfellow et al., 2016). A convolution block contains a padded $3 \times 3$ convolutional layer, followed by a batch normalization layer and a rectified linear unit (ReLU) activation function, which is given by $\mathrm{ReLU}\colon \mathbb{R} \to \mathbb{R}_0^+ = [0, \infty)$, $\mathrm{ReLU}(x) = \max(x, 0)$. The convolution blocks of the first downsampling step produce $n_\mathrm{f} \in \mathbb{N}$ feature channels, and each further downsampling step doubles the number of feature channels. The bottleneck at the fifth layer consists of two such convolution blocks. For each of the subsequent upsampling steps, the spatial resolution of the feature channels is increased using bilinear interpolation (Dumitrescu and Boiangiu, 2019). The resulting channels are then concatenated with the feature channels of the corresponding level of the downsampling path. Two convolutional blocks following the concatenation of those feature channels divide the number of feature channels by three and complete the upsampling step. Here, the number of feature channels is divided by three since the concatenated channels consist of the channels of the downsampling path and twice as many interpolated channels from the level below. After the final upsampling step, the sigmoid activation function $\sigma\colon \mathbb{R} \to [0, 1]$, given by

$$\sigma(x) = \frac{1}{1 + e^{-x}},$$

is applied to the output.

Compared to the original U-net architecture presented in Ronneberger et al. (2015), we use padded convolutional layers with a padding of 1 in each step, preserving the image size throughout the network. Since the network input is periodic in the second component, we use periodic padding (Schubert et al., 2019) in this direction, while the padding in the direction of the first component is reflective. Moreover, the spatial resolution of the features is upsampled using bilinear interpolation instead of an up-convolution. Furthermore, the network does not include a final $1 \times 1$ convolution. Instead, the output of the final convolution block is already a single-channel map to which the sigmoid activation function is applied directly. A sketch of the network architecture described above is shown in Fig. 4.

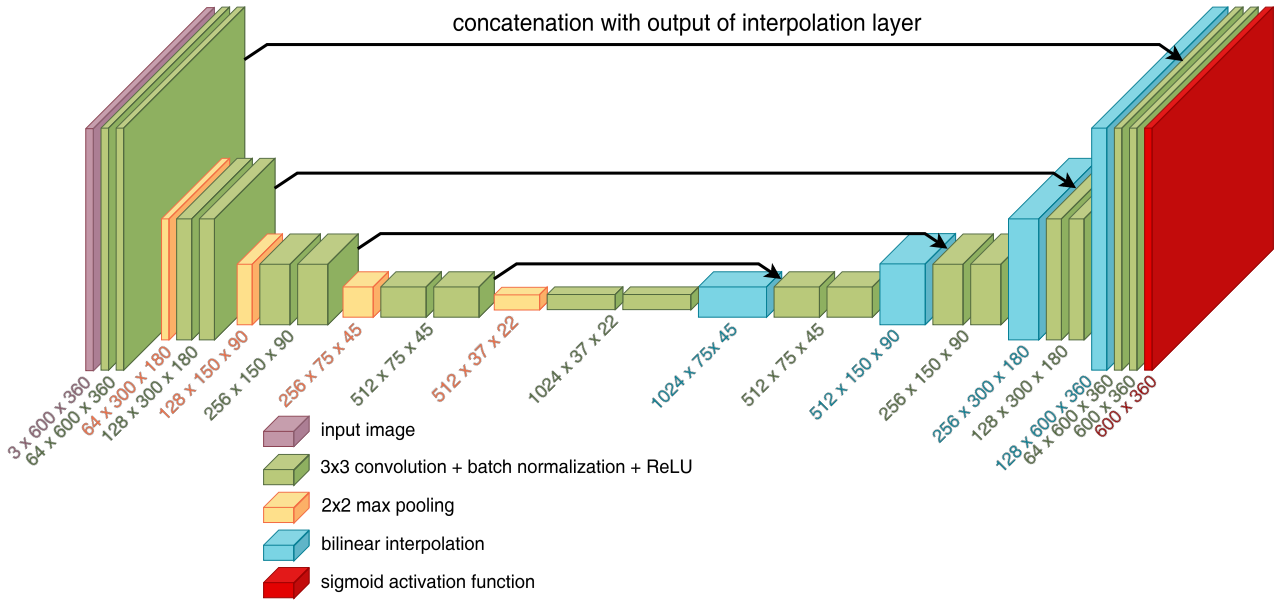


Figure 4: Schematic representation of the network architecture with $n_\mathrm{f} = 64$ feature channels in the first downsampling step. Below each layer, the respective output dimensions are given.

The network outputs are images of the same spatial dimensions as each of the input channels. More specifically, the network architecture stated above with parameter constellation $\theta \in \Theta \subset \mathbb{R}^q$ is given by a function $f_\theta \colon \mathbb{R}^{k \times h \times w} \to [0,1]^{h \times w}$, with spatial dimensions $h, w \in \mathbb{N}$ and the number of input channels $k = 3$. Here, $\theta$ denotes the vector of trainable parameters of the network, and $\Theta \subset \mathbb{R}^q$ is the network's parameter space for some $q \in \mathbb{N}$ depending on the number of feature channels $n_\mathrm{f}$. Each input channel corresponds to a single-channel image of a radar moment, as described in Sect. 2.1, allowing the network to combine information about different radar moments to detect non-meteorological echoes.

The goal is then to perform supervised training of the network to assign binary labels to each pixel $(i, j) \in \Omega$, describing which pixels correspond to meteorological and non-meteorological echoes, respectively. This training process requires pixel-wise labeled training data, that is, each pixel $(i, j) \in \Omega$ must be assigned a ground truth class label $c_{i,j} \in \{0, 1\}$. However, mixed radar images containing both meteorological and non-meteorological echoes lack pixel-wise ground truth labels, and generating such labels by hand is very costly. Therefore, we propose a method to generate synthetic multi-channel images and corresponding pixel-wise ground truth labels for training the network.

## 2.4 Generation of synthetic training data

As outlined in Sect. 2.3, the network architecture considered in this paper can utilize radar images as input, where each channel corresponds to a single-channel image of a specific radar moment. The training samples consist of such radar images and corresponding binary ground truth images indicating whether each pixel corresponds to a meteorological or a non-meteorological echo. We generate synthetic radar images by combining radar images corresponding to winter sweeps (featuring meteorological echoes) and radar images corresponding to cluttered summer sweeps (featuring non-meteorological echoes). For brevity, we refer to these as winter radar images and cluttered summer radar images, respectively. This approach is motivated by the observation that the horizontal reflectivity (DBZH) values in mixed summer radar images often have the appearance of a combination of (scaled) DBZH values from winter and from cluttered summer radar images, as illustrated in Fig. 1b. In this way, we can generate labeled training samples where the contributions of meteorological and non-meteorological echoes are explicitly known, addressing the challenge of difficult labeling in mixed radar images. However, the meteorological echoes in the synthetic radar images are exclusively those of the winter radar images. Consequently, precipitation structures that occur in summer only are not represented in the training data.

More specifically, we generate a realistic mixed (synthetic) DBZH image by combining two DBZH images $A^{\mathrm{H,w}}, A^{\mathrm{H,s}} \in \mathbb{R}^{n_\mathrm{r} \times n_\mathrm{a}}$ from a winter and a cluttered summer radar image in the following manner. Since the horizontal reflectivity is measured in decibels, which is a logarithmic unit of measurement, we first transform the entries of the DBZH images into linear scale using the transformation $\Phi\colon \mathbb{R} \to \mathbb{R}^+ = (0, \infty)$ (Ryzhkov and Zrnić, 2019), which is given by

$$\Phi(x) = 10^{x/10}.$$

The transformed reflectivity values are scaled by suitable weighting factors $\lambda_\mathrm{w} > 0$ and $\lambda_\mathrm{s} > 0$ for the DBZH images from the winter and the cluttered summer radar image, respectively. For each pixel, the scaled reflectivity values are summed up and transformed back into decibels to obtain the synthetic DBZH image $A^{\mathrm{H,m}} \in \mathbb{R}^{n_\mathrm{r} \times n_\mathrm{a}}$ with entries

$$A^{\mathrm{H,m}}_{i,j} = \begin{cases} \Phi^{-1}(\lambda_\mathrm{w}\Phi(A^{\mathrm{H,w}}_{i,j}) + \lambda_\mathrm{s}\Phi(A^{\mathrm{H,s}}_{i,j})), & \text{if } A^{\mathrm{H,w}}_{i,j} > -40\,\mathrm{dBZ} \text{ or } A^{\mathrm{H,s}}_{i,j} > -40\,\mathrm{dBZ}, \\ -40, & \text{otherwise}, \end{cases}$$

for each $(i, j) \in \Omega$.

Similarly to the placeholder values described in Sect. 2.2, we assign the entry of the synthetic DBZH image $A^{\mathrm{H,m}}_{i,j}$ a placeholder value $p = -40\,\mathrm{dBZ}$ if the pixel is excluded for both radar images. To find appropriate weighting factors $\lambda_\mathrm{w}$ and $\lambda_\mathrm{s}$, we consider the difference in horizontal reflectivity values observed in Fig. 1. Since the range of the horizontal reflectivity values corresponding to non-meteorological echoes is almost identical for cluttered summer radar images (Fig. 1c) and mixed summer radar images (Fig. 1b), the choice of $\lambda_\mathrm{s} = 1$ is suitable. However, the horizontal reflectivity values corresponding to meteorological echoes are substantially lower for winter radar images (Fig. 1a) compared to mixed summer radar images. Since the resulting synthetic DBZH image should resemble the typical DBZH images from mixed summer radar images, a weighting factor $\lambda_\mathrm{w} > 1$ is necessary. In order to determine a suitable choice, four representative mixed summer radar images were chosen. For each of them, a winter radar image with precipitation areas of similar size was assigned. Then, for different weighting factors $\lambda_\mathrm{w} \in \{1.5, 3, 5, 8, 15\}$, we compared the DBZH values of manually sampled pixels within similar meteorological structures of both images. As a result, a weighting factor $\lambda_\mathrm{w}$ between 5 and 8 was deemed appropriate. Note that the weighting factors determined in this way are specific

to the considered time periods, though the considered range for the weighting factor allows for calibration differences of up to 3 dB. Figure 5 shows how a synthetic DBZH image is generated by combining a winter and a cluttered summer radar image using a weighting factor $\lambda_{\mathrm{w}} = 6.5$. The resulting synthetic DBZH image (Fig. 5c) is then compared with one from a mixed summer radar image (Fig. 5d).

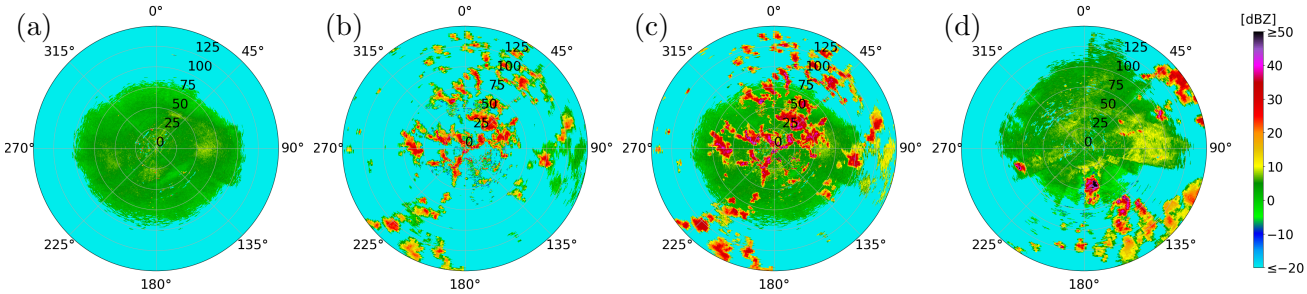


Figure 5: Comparison of Cartesian representations of different DBZH images: (a) cluttered summer radar image, (b) winter radar image, (c) their corresponding synthetic radar image with weighting factors $\lambda_{\mathrm{w}} = 6.5$, and $\lambda_{\mathrm{s}} = 1$ and (d) mixed summer radar image. The radar images shown were measured (a) on June 18, 2022, at 10:00 UTC in Neuheilenbach, (b) on January 9, 2025, at 17:00 UTC in Borkum and (d) on June 16, 2022, at 09:55 UTC in Memmingen.

The two chosen winter and cluttered summer radar images are then also used to generate synthetic single-channel images for the other radar moments considered in Sect. 2.1. In addition to the horizontal reflectivity, which is proportional to the particle concentration and the particle's horizontal cross-section within the radar gate, the differential reflectivity (ZDR) primarily reflects the mean axis ratio of the particles (Straka et al., 2000). Since differential reflectivity is defined as the difference between horizontal and vertical reflectivity in decibels, synthetic values can be generated by taking the difference between synthetic horizontal and vertical reflectivity values. Similarly to the synthetic DBZH image $A^{\mathrm{H,m}}$, a synthetic DBZV image $A^{\mathrm{V,m}} \in \mathbb{R}^{n_{\mathrm{r}} \times n_{\mathrm{a}}}$ is also generated. Then, the corresponding synthetic ZDR image is denoted by $A^{\mathrm{ZDR,m}} \in \mathbb{R}^{n_{\mathrm{r}} \times n_{\mathrm{a}}}$, and its entries are given by

$$A_{i,j}^{\mathrm{ZDR,m}} = \begin{cases} A_{i,j}^{\mathrm{H,m}} - A_{i,j}^{\mathrm{V,m}}, & \text{if } A_{i,j}^{\mathrm{H,m}} > -40 \text{ and } A_{i,j}^{\mathrm{V,m}} > -40, \\ 20, & \text{if } A_{i,j}^{\mathrm{H,m}} = -40 \text{ or } A_{i,j}^{\mathrm{V,m}} = -40, \end{cases}$$

for each $(i, j) \in \Omega$, see Fig. 6. Following Sect. 2.2, a placeholder value of $p = 20$ dB is assigned to all entries of the synthetic ZDR image for which at least one of the corresponding synthetic DBZH and DBZV values is a placeholder value.

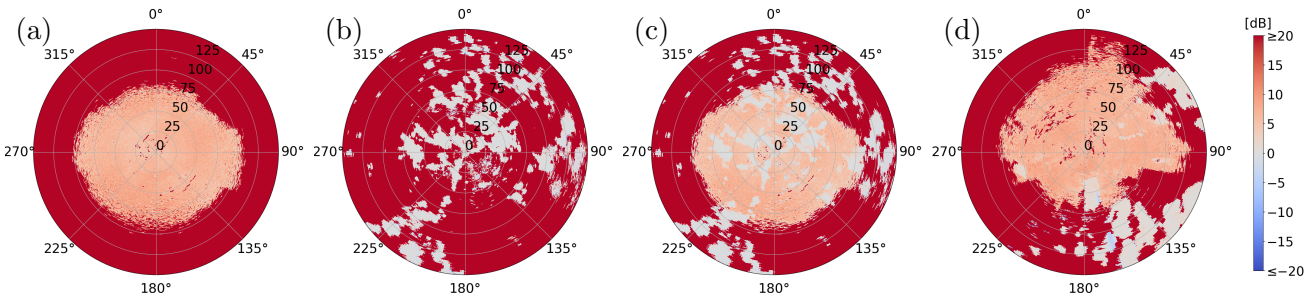


Figure 6: Comparison of Cartesian representations of different ZDR images: (a) cluttered summer radar image, (b) winter radar image, (c) their corresponding synthetic radar image, and (d) mixed summer radar image. The radar images shown were measured at the times and the radar stations mentioned in Fig. 5.

In addition to describing the size and shape of particles using DBZH and ZDR, we also use the depolarization ratio UDR to quantify the homogeneity of the detected particles' shapes within a radar gate. To generate a synthetic UDR image $A^{\mathrm{UDR,m}} \in (-\infty, 0]^{n_\mathrm{r} \times n_\mathrm{a}}$, we use the property that pixels $(i, j) \in \Omega$ included in the winter radar image are assumed to correspond to meteorological echoes. The corresponding entries from the synthetic UDR image are assigned the depolarization ratio values $A^{\mathrm{UDR,w}} \in (-\infty, 0]^{n_\mathrm{r} \times n_\mathrm{a}}$ from the winter radar image. Pixels excluded from the winter radar image are assigned the values $A^{\mathrm{UDR,s}} \in (-\infty, 0]^{n_\mathrm{r} \times n_\mathrm{a}}$ from the cluttered summer radar image if they are included in this radar image. Otherwise, they are assigned a placeholder value. More precisely, the entries of $A^{\mathrm{UDR,m}}$ are given by

$$A_{i,j}^{\mathrm{UDR,m}} = \begin{cases} A_{i,j}^{\mathrm{UDR,w}}, & \text{if } A^{\mathrm{H,w}} > -40\,\mathrm{dBZ}, \\ A_{i,j}^{\mathrm{UDR,s}}, & \text{if } A^{\mathrm{H,w}} = -40\,\mathrm{dBZ} \text{ and } A^{\mathrm{H,s}} > -40\,\mathrm{dBZ}, \\ 0, & \text{otherwise}, \end{cases}$$

for each $(i, j) \in \Omega$, see Fig. 7. For pixels included in both radar images, the synthetic DBZH value combines both DBZH values. However, the synthetic UDR value for these pixels is taken from the winter radar image alone, since there is no physically obvious way of combining the two depolarization ratios.

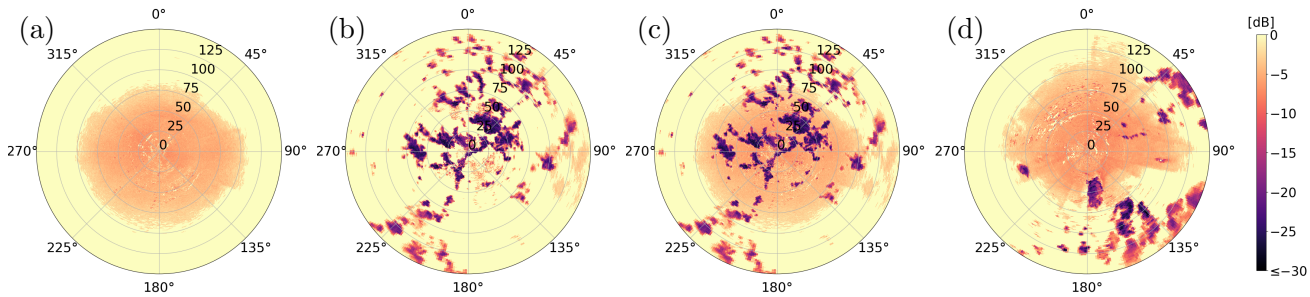


Figure 7: Comparison of Cartesian representations of different UDR images: (a) cluttered summer radar image, (b) winter radar image, (c) their corresponding synthetic radar image, and (d) mixed summer radar image. The radar images shown were measured at the times and the radar stations mentioned in Fig. 5.

Alongside the synthetic radar image $A^{\mathrm{m}} = (A^{\mathrm{H,m}}, A^{\mathrm{ZDR,m}}, A^{\mathrm{UDR,m}}) \in \mathbb{R}^{3 \times n_\mathrm{r} \times n_\mathrm{a}}$, we generate a corresponding ground truth image $G \in \{0, 1\}^{n_\mathrm{r} \times n_\mathrm{a}}$. Its values indicate which pixels of the synthetic radar image correspond to meteorological ($G_{i,j} = 0$) and non-meteorological echoes ($G_{i,j} = 1$), respectively. Pixels included in the winter radar image $A^{\mathrm{H,w}}$ are assumed to correspond to meteorological echoes, as described in Sect. 2.2. Since the synthetic image is generated in such a way that these pixels also correspond to meteorological echoes in the synthetic image, they are assigned a ground truth label of 0. In particular, pixels that are also included in the cluttered summer radar image, and whose synthetic reflectivity values therefore contain both a meteorological and a non-meteorological contribution, are labeled as meteorological. In this way, every pixel corresponding to meteorological echoes is labeled as such. All of the other pixels, i.e., those excluded from the winter radar image, are labeled 1. More precisely, for each pixel $(i, j) \in \Omega$, the ground truth label $G_{i,j}$ is given by

$$G_{i,j} = \begin{cases} 0, & \text{if } A^{\mathrm{H,w}} > -40\,\mathrm{dBZ}, \\ 1, & \text{otherwise}. \end{cases}$$

Generating synthetic radar images in this way provides mixed radar images with pixel-wise ground truth labeling, assigning each pixel as either meteorological or non-meteorological echoes.

This ground truth allows the three synthetic channels to serve as training samples for the neural network.

## 2.5 Network training

The available winter dataset $\mathcal{D}^{\mathrm{w}} = \{A_n^{\mathrm{w}}\}_{n=1}^{180} \subset D = \mathbb{R}^{3\times n_{\mathrm{r}}\times n_{\mathrm{a}}}$ consists of 180 radar images $A_n^{\mathrm{w}} = (A_n^{\mathrm{H,w}}, A_n^{\mathrm{V,w}}, A_n^{\mathrm{UDR,w}})$, where the radar moments present in the data are horizontal reflectivity, vertical reflectivity, and the depolarization ratio. Similarly, the available cluttered summer dataset is given by $\mathcal{D}^{\mathrm{s}} = \{A_n^{\mathrm{s}}\}_{n=1}^{180} \subset D$. We partition each of these datasets into two disjoint subsets. One subset of radar images is used for generating the training samples, while the other subset is reserved for generating the validation samples, which are used for model selection, see Sect. 2.6. For this purpose, we partition the index set $\mathcal{I} = \{1, \ldots, 180\}$ into two disjoint subsets $\mathcal{I}_{\mathrm{train}}, \mathcal{I}_{\mathrm{val}} \subset \mathcal{I}$, i.e.,

$$\mathcal{I}_{\mathrm{train}} \cup \mathcal{I}_{\mathrm{val}} = \mathcal{I} \quad \text{and} \quad \mathcal{I}_{\mathrm{train}} \cap \mathcal{I}_{\mathrm{val}} = \emptyset.$$

The index sets $\mathcal{I}_{\mathrm{train}}, \mathcal{I}_{\mathrm{val}} \subset \mathcal{I}$ are chosen so that all winter and cluttered summer radar images from a single radar station (radar Essen) are used for the generation of validation samples, while the radar images from other radar stations are used for the generation of training samples.

We use the methodology described in Sect. 2.4 to generate synthetic training samples by combining radar images from the winter and cluttered summer datasets. Each synthetic training sample consists of a synthetic radar image $A \in D$ and a corresponding ground truth image $G \in \{0,1\}^{n_{\mathrm{r}}\times n_{\mathrm{a}}}$. The synthetic radar image comprises the radar moments of horizontal reflectivity, differential reflectivity, and the depolarization ratio, i.e., $A = (A^{\mathrm{H}}, A^{\mathrm{ZDR}}, A^{\mathrm{UDR}})$. Before generating the synthetic training samples, we apply various data augmentation steps to increase their diversity. More precisely, as described in Sect. 2.4, the DBZH and DBZV images from the winter and cluttered summer datasets are scaled by reflectivity weighting factors $\lambda_{\mathrm{w}}, \lambda_{\mathrm{s}} > 0$, respectively. In addition, we apply a rotation and a possible inversion of orientation.

Specifically, we introduce a parametric mapping $g_{\psi,\psi'} \colon D \times D \to D \times \{0,1\}^{n_{\mathrm{r}}\times n_{\mathrm{a}}}$ that combines two radar images to generate a synthetic radar image and a corresponding ground truth image. The space of admissible parameters for data augmentation is given by $\Psi = \mathbb{R}^+ \times \{1, \ldots, n_{\mathrm{a}}\} \times \{-1, 1\}$. Each parameter vector $\psi = (\lambda, \varphi, o) \in \Psi$ contains the reflectivity weighting factor $\lambda > 0$, the rotation angle $\varphi \in \{1, \ldots, n_{\mathrm{a}}\}$, and the orientation flag $o \in \{-1, 1\}$. Here, a rotation angle of $\varphi \in \{1, \ldots, n_{\mathrm{a}}\}$ indicates a rotation by $\varphi$ rays. Hence, a shift of $\varphi$ rows is applied to the radar image. An orientation flag of $o = -1$ indicates an inverted orientation. In that case, the radar image is flipped along the $x$-axis. To distinguish the transformations we apply to winter and cluttered summer radar images, we introduce restricted admissible parameter spaces $\Psi^{\mathrm{w}}, \Psi^{\mathrm{s}} \subset \Psi$. They only differ in the admissible range of the reflectivity weighting factor, which is given by the interval $[5, 8]$ for $\Psi^{\mathrm{w}}$ and by $[0.9, 1.1]$ for $\Psi^{\mathrm{s}}$. The set of synthetic training samples generated by combining two radar images is then denoted by

$$\mathcal{D}_{\mathrm{train}}^{\mathrm{m}} = \left\{ g_{\psi_{\mathrm{w}},\psi_{\mathrm{s}}}(A_{n_{\mathrm{w}}}^{\mathrm{w}}, A_{n_{\mathrm{s}}}^{\mathrm{s}}) : \psi_{\mathrm{w}} \in \Psi^{\mathrm{w}}, \psi_{\mathrm{s}} \in \Psi^{\mathrm{s}}, n_{\mathrm{w}}, n_{\mathrm{s}} \in \mathcal{I}_{\mathrm{train}} \right\}.$$

Note that, in order to make use of as many different combinations as possible, the indices $n_{\mathrm{w}}, n_{\mathrm{s}} \in \mathcal{I}_{\mathrm{train}}$ of the winter and the cluttered summer radar image of a training sample are not matched by radar station or measurement time. In particular, radar images from different radar stations are combined for the generation of synthetic radar images.

These synthetic training samples feature both meteorological and non-meteorological echoes. However, in the training process, the network should be provided with training samples from various scenarios. Since the radar images from the winter and cluttered summer datasets, $\mathcal{D}^{\mathrm{w}}$

and $\mathcal{D}^{\mathrm{s}}$, are used to represent meteorological and non-meteorological echoes, respectively, it is possible to generate synthetic radar images that feature only one type of echo. To this end, we introduce a radar image $A^{\mathrm{exc}} \in D$ in which all pixels are excluded, i.e., all of its entries are assigned the corresponding placeholder values. A synthetic radar image that features only meteorological echoes is then generated by combining a winter radar image and the radar image $A^{\mathrm{exc}}$. In particular, the set of training samples featuring only meteorological echoes is given by

$$\mathcal{D}^{\mathrm{w}}_{\mathrm{train}} = \left\{ g_{\psi_{\mathrm{w}},\psi}(A^{\mathrm{w}}_{n_{\mathrm{w}}}, A^{\mathrm{exc}}) : \psi_{\mathrm{w}} \in \Psi^{\mathrm{w}}, \psi \in \Psi, n_{\mathrm{w}} \in \mathcal{I}_{\mathrm{train}} \right\}.$$

Analogously, the set of synthetic radar images featuring only non-meteorological echoes is given by

$$\mathcal{D}^{\mathrm{s}}_{\mathrm{train}} = \left\{ g_{\psi,\psi_{\mathrm{s}}}(A^{\mathrm{exc}}, A^{\mathrm{s}}_{n_{\mathrm{s}}}) : \psi \in \Psi, \psi_{\mathrm{s}} \in \Psi^{\mathrm{s}}, n_{\mathrm{s}} \in \mathcal{I}_{\mathrm{train}} \right\}.$$

Note that the validation dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$, comprising radar images featuring both meteorological and non-meteorological echoes, and the validation datasets $\mathcal{D}^{\mathrm{w}}_{\mathrm{val}}$ and $\mathcal{D}^{\mathrm{s}}_{\mathrm{val}}$, comprising radar images featuring only one type of echo, are generated analogously using the index set $\mathcal{I}_{\mathrm{val}}$ instead of $\mathcal{I}_{\mathrm{train}}$.

As stated in Sect. 2.3, the network can be described by its vector of trainable parameters $\theta \in \Theta$ using a function $f_\theta \colon \mathbb{R}^{k \times h \times w} \to [0,1]^{h \times w}$. Here, $h = n_{\mathrm{r}}$ and $w = n_{\mathrm{a}}$ denote the image dimensions of a single-channel image, and $k = 3$ is the number of input channels. Then, for a given radar image $A \in D$, the output $f_\theta(A)$ of the network represents the matrix of probabilities $(f_\theta(A))_{i,j}$ that a pixel $(i,j)$ corresponds to a non-meteorological echo for each $(i,j) \in \Omega$.

The training process of the network involves iterating over so-called mini-batches, and the trainable parameters of the network are adapted at each iteration. More precisely, in each training step, a mini-batch $B = \{(A_n, G_n)\}_{n=1}^4$ is built by generating a total of four synthetic radar images and corresponding ground truth images. Each training sample is chosen as follows: With a probability of 0.75, it is uniformly sampled from the training dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{train}}$ and therefore features both meteorological and non-meteorological echoes. Otherwise, we either sample it from $\mathcal{D}^{\mathrm{w}}_{\mathrm{train}}$ or $\mathcal{D}^{\mathrm{s}}_{\mathrm{train}}$, with a probability of 0.125 for each case. The mixed radar images are sampled most frequently since only they contain boundaries between meteorological and non-meteorological echoes, which are the most difficult regions to classify. The winter and cluttered summer samples ensure that the network also handles radar images featuring only one type of echo. For each training sample $(A_n, G_n)$ of the mini-batch $B$, with $n \in \{1,2,3,4\}$, the network output $f_\theta(A_n)$ is computed and compared to the ground truth image $G_n$ using the binary cross-entropy loss, which is a measure of how well the network outputs match the ground truth images. More specifically, for any batch $B = \{(A_n, G_n)\}_{n=1}^4$ and vector of trainable network parameters $\theta \in \Theta$, the cross-entropy loss $L_\theta(B)$ of the batch is given by

$$L_\theta(B) = -\frac{1}{4} \sum_{n=1}^{4} \sum_{(i,j) \in \Omega} \ell(f_\theta(A_n)_{i,j}, (G_n)_{i,j}),$$

where $\ell(y,x) = x \log(y) + (1-x)\log(1-y)$, with $f_\theta(A_n)_{i,j}$ and $(G_n)_{i,j}$ denoting the entries of $f_\theta(A_n)$ and $G_n$, respectively. After computing the cross-entropy loss $L_\theta(B)$, its gradient is computed with respect to the vector of trainable network parameters $\theta$. Based on that, this parameter vector $\theta$ is then adjusted using the Adam optimizer (Kingma and Ba, 2017) with a learning rate of $10^{-3}$. The network was implemented and trained using PyTorch 2.8.0 (Ansel et al., 2024). The batch size, the sampling probabilities, and the learning rate were fixed before training and are not subject to the model selection described in Sect. 2.6. An overview of the hyperparameters of the network is shown in Table 1.

Table 1: Hyperparameters of the neural network and its training.

| Hyperparameter | Value |
|---|---|
| number of feature channels $n_\mathrm{f}$ | $n_\mathrm{f} \in \{8, 16, 32, 64\}$ |
| mini-batches per epoch | 42 |
| batch size | 4 |
| sampling probabilities for $\mathcal{D}^\mathrm{m}_\mathrm{train}$, $\mathcal{D}^\mathrm{w}_\mathrm{train}$, $\mathcal{D}^\mathrm{s}_\mathrm{train}$ | 0.75, 0.125, 0.125 |
| ranges of the reflectivity weighting factors $\lambda_\mathrm{w}$, $\lambda_\mathrm{s}$ | $[5, 8]$, $[0.9, 1.1]$ |
| optimizer | Adam |
| learning rate | $10^{-3}$ |
| loss function | binary cross-entropy |

## 2.6 Model selection and early stopping

The number $n_\mathrm{f}$ of feature channels, see Sect. 2.3, determines the number of trainable parameters of the network. If it is too small, the network cannot learn the features that distinguish meteorological from non-meteorological echoes; if it is too large, the network can overfit, i.e., it may adapt to details of the synthetic training data that do not carry over to measured radar images. Similarly, the network is expected to fit the training data more closely with each training step, which can also lead to overfitting. Since suitable values are difficult to determine a priori, the number $n_\mathrm{f}$ of feature channels and the length of training are selected based on the validation datasets described in Sect. 2.5. More precisely, a fixed validation set $V = \{(A^\mathrm{val}_n, G^\mathrm{val}_n)\}_{n=1}^N \subset \mathcal{D}^\mathrm{m}_\mathrm{val} \cup \mathcal{D}^\mathrm{w}_\mathrm{val} \cup \mathcal{D}^\mathrm{s}_\mathrm{val}$ consisting of $N = 144$ validation samples is generated once before the training process. The training samples are generated in each training step as described in Sect. 2.5. Then, for each $n_\mathrm{f} \in \{8, 16, 32, 64\}$, one network is trained as described above for up to 100 epochs consisting of 42 training steps each. For each $e \in \{1, \ldots, 100\}$, its validation loss after $e$ epochs is given by

$$L^\mathrm{val}_{n_\mathrm{f},e} = -\frac{1}{N} \sum_{n=1}^{N} \sum_{(i,j)\in\Omega} \ell(f_{\theta_{n_\mathrm{f},e}}(A^\mathrm{val}_n)_{i,j}, (G^\mathrm{val}_n)_{i,j}),$$

where $\theta_{n_\mathrm{f},e}$ denotes the network's trainable parameters after $e$ epochs. To prevent overfitting, the training is stopped early once the validation loss has not decreased over three consecutive intervals of five epochs, i.e., after the smallest epoch $e \in \{20, 25, \ldots, 100\}$ for which $L^\mathrm{val}_{n_\mathrm{f},e-15} \leq L^\mathrm{val}_{n_\mathrm{f},e-10} \leq L^\mathrm{val}_{n_\mathrm{f},e-5} \leq L^\mathrm{val}_{n_\mathrm{f},e}$. If there is no such epoch, the training ends after 100 epochs. The last epoch is denoted by $e^\mathrm{end}_{n_\mathrm{f}}$.

The number of feature channels and training epochs of the network are selected based on the moving average of the validation loss over the set $E_{n_\mathrm{f},e} = \{e-2, \ldots, e+2\} \cap \{1, \ldots, e^\mathrm{end}_{n_\mathrm{f}}\}$ of up to five epochs surrounding $e \in \{1, \ldots, e^\mathrm{end}_{n_\mathrm{f}}\}$, given by

$$\bar{L}^\mathrm{val}_{n_\mathrm{f},e} = \frac{1}{|E_{n_\mathrm{f},e}|} \sum_{k\in E_{n_\mathrm{f},e}} L^\mathrm{val}_{n_\mathrm{f},k}.$$

This moving average is considered in order to reduce the influence of fluctuations between single epochs, which can be quite large due to the small validation set. For the first two and the last two epochs, the set $E_{n_\mathrm{f},e}$ contains only three or four epochs, so that the last epochs are not excluded from the selection. Among all pairs $(n_\mathrm{f}, e)$ with $n_\mathrm{f} \in \{8, 16, 32, 64\}$ and $e \in \{1, \ldots, e^\mathrm{end}_{n_\mathrm{f}}\}$, the selected number of feature channels $n^*_\mathrm{f}$ and the selected epoch $e^*$ are then given by

$$(n^*_\mathrm{f}, e^*) = \underset{\substack{n_\mathrm{f}\in\{8,16,32,64\} \\ e\in\{1,\ldots,e^\mathrm{end}_{n_\mathrm{f}}\}}}{\operatorname{argmin}} \bar{L}^\mathrm{val}_{n_\mathrm{f},e},$$

and the trainable parameters of the selected network are denoted by $\theta^* = \theta_{n_\mathrm{f}^*, e^*}$.

## 2.7 Metrics for quantitative evaluation

To quantitatively evaluate the segmentation performance, we compare the network output against corresponding ground truth. For that purpose, pixels $(i, j) \in \Omega$ with a horizontal reflectivity value of $A^\mathrm{H}_{i,j} = -40\,\mathrm{dBZ}$ are excluded from the evaluation, as they conclusively do not correspond to meteorological echoes. Furthermore, the physical area of a radar gate $g_{i,j}$ increases linearly with its distance from the radar. Therefore, to obtain metrics with a meaningful physical interpretation regarding the spatial extent of the echoes, we introduce a weight matrix $W = (W_{i,j}) \in \mathbb{R}^{n_\mathrm{r} \times n_\mathrm{a}}$. For each pixel $(i, j) \in \Omega$, its weight $W_{i,j}$ is given by

$$W_{i,j} = \begin{cases} a_i, & \text{if } A^\mathrm{H}_{i,j} > -40\,\mathrm{dBZ}, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $a_i = \frac{\pi}{n_\mathrm{a}}(2r_i - w)w$ denotes the physical area of radar gate $g_{i,j}$, where $w = 250$ is the range resolution in meters, and $r_i = iw$ is the outer radius of the radar gate. Using the weights of the individual pixels, we define the weighted area $\mathcal{A}(\Omega')$ of any subset of pixels $\Omega' \subset \Omega$ as the sum of the weights of the pixels in that set, given by

$$\mathcal{A}(\Omega', W) = \sum_{(i,j) \in \Omega'} W_{i,j}.$$

Using a radar image $A \in D = \mathbb{R}^{3 \times n_\mathrm{r} \times n_\mathrm{a}}$ as input, the network output $f_{\theta^*}(A) \in [0, 1]^{n_\mathrm{r} \times n_\mathrm{a}}$ is transformed into a binary prediction $P = (P_{i,j}) \in \{0, 1\}^{n_\mathrm{r} \times n_\mathrm{a}}$ by thresholding at 0.5. Specifically, the prediction $P_{i,j}$ at pixel $(i, j) \in \Omega$ is given by

$$P_{i,j} = \begin{cases} 0, & \text{if } (f_{\theta^*}(A))_{i,j} < 0.5, \\ 1, & \text{if } (f_{\theta^*}(A))_{i,j} \geq 0.5, \end{cases}$$

where prediction $P_{i,j} = 0$ represents meteorological echoes, and $P_{i,j} = 1$ represents non-meteorological echoes.

To quantify segmentation performance, we compute the area-weighted intersection over union (IOU), also referred to as the critical success index (Mbizvo and Larner, 2024). For class $c \in \{0, 1\}$, the class-specific IOU is defined as the ratio of the area of the intersection of pixels classified as class $c$ in the prediction and pixels labeled as class $c$ in the ground truth, to the area of their union, given by

$$\mathrm{IOU}^c(G, P, W) = \frac{\mathcal{A}(\Omega|_{P=c} \cap \Omega|_{G=c}, W)}{\mathcal{A}(\Omega|_{P=c} \cup \Omega|_{G=c}, W)}.$$

Here, the set of pixels assigned to class $c$ by the prediction $P$ and the set of pixels labeled as class $c$ in the ground truth $G$ are given by

$$\Omega|_{P=c} = \{(i, j) \in \Omega : P_{i,j} = c\} \quad \text{and} \quad \Omega|_{G=c} = \{(i, j) \in \Omega : G_{i,j} = c\},$$

respectively. We then obtain a single performance metric for the entire radar image by computing the combined IOU as the area-weighted sum of the class-specific IOUs, given by

$$\mathrm{IOU}(G, P, W) = \sum_{c \in \{0,1\}} \frac{\mathcal{A}(\Omega|_{G=c})}{\mathcal{A}(\Omega)} \mathrm{IOU}^c(G, P, W).$$

While the area-weighted IOU is an effective aggregate metric for overall segmentation quality, it does not distinguish between incorrectly classifying pixels as meteorological and incorrectly classifying pixels as non-meteorological echoes. Therefore, it is helpful to quantify the portion of correct predictions within specific regions, e.g., regions labeled as class $c = 0$ in the ground truth. To this end, we define the area-weighted accuracy on a subset $\Omega' \subseteq \Omega$. This metric calculates the portion of the weighted area within $\Omega'$ that is correctly classified and is given by

$$\mathrm{ACC}(G, P, W; \Omega') = \frac{\mathcal{A}(\Omega' \cap \Omega|_{P=G})}{\mathcal{A}(\Omega')},$$

where $\Omega|_{P=G} = \{(i, j) \in \Omega : P_{i,j} = G_{i,j}\}$ denotes the set of all correctly classified pixels. Setting $\Omega' = \Omega$ yields the area-weighted accuracy for the entire radar image.

This general formulation allows us to derive the class-specific precision and recall, which provide insight into specific error types. The precision $\mathrm{PRE}^c$ corresponds to the accuracy evaluated on $\Omega|_{P=c}$, i.e., the set of pixels classified as class $c$. It quantifies the correctness of the prediction, where a low precision implies a high rate of false positives for class $c$, i.e., a large portion of pixels classified as class $c$ are incorrectly classified. Conversely, the recall $\mathrm{REC}^c$ corresponds to the accuracy evaluated on $\Omega|_{G=c}$, i.e., the set of pixels labeled as class $c$ in the ground truth. It measures the completeness of the detection, where a low recall implies a high rate of false negatives, i.e., a large number of pixels with ground truth label $c$ are incorrectly classified. These metrics are given by

$$\mathrm{PRE}^c(G, P, W) = \mathrm{ACC}(G, P, W; \Omega|_{P=c}) \quad \text{and} \quad \mathrm{REC}^c(G, P, W) = \mathrm{ACC}(G, P, W; \Omega|_{G=c}).$$

While false positives for meteorological echoes (lower $\mathrm{REC}^1$ and $\mathrm{PRE}^0$) introduce undesirable noise, missing weather events (lower $\mathrm{REC}^0$ and $\mathrm{PRE}^1$) constitute a severe loss of information. From a meteorological perspective, a high recall $\mathrm{REC}^0$ for meteorological echoes is essential to ensure that weather events are correctly identified. This is particularly important for meteorological echoes with high reflectivity values, as they correspond to severe weather events. Therefore, we also investigate the network performance with respect to the horizontal reflectivity values $A^{\mathrm{H}}$. For this, we evaluate the segmentation performance on subsets of pixels with horizontal reflectivity values $A^{\mathrm{H}}_{i,j}$ in the semi-open interval $[z, z + 5)$ for $z \in \mathbb{R}$, given by

$$\Omega|_{A^{\mathrm{H}} \in [z, z+5)} = \{(i, j) \in \Omega : A^{\mathrm{H}}_{i,j} \in [z, z + 5)\}.$$

Similarly, we analyze the segmentation performance with respect to the distance from the radar, i.e., on subsets of pixels

$$\Omega|_{i \in [r, r+19]} = \{(i, j) \in \Omega : r \leq i \leq r + 19\},$$

for range indices $r \in \{1, \ldots, n_{\mathrm{r}} - 19\}$. This is motivated by the fact that non-meteorological echoes in winter radar images, which are rare but can appear close to the radar station, as shown in Fig. 3d, result in incorrect ground truth labels in the synthetically generated training data.

Using these subsets of pixels, we compute the area-weighted accuracy depending on reflectivity and distance from the radar, given by

$$\mathrm{ACC}^{\mathrm{H}}(G, P, W; z) = \mathrm{ACC}(G, P, W; \Omega|_{A^{\mathrm{H}} \in [z, z+5)}),$$
$$\mathrm{ACC}^{\mathrm{D}}(G, P, W; r) = \mathrm{ACC}(G, P, W; \Omega|_{i \in [r, r+19]}),$$

respectively, and the recall for class $c \in \{0,1\}$ depending on reflectivity and distance from the radar, given by

$$\mathrm{REC}^{c,\mathrm{H}}(G,P,W;z) = \mathrm{ACC}(G,P,W;\Omega|_{G=c} \cap \Omega|_{A^{\mathrm{H}}\in[z,z+5)}),$$
$$\mathrm{REC}^{c,\mathrm{D}}(G,P,W;r) = \mathrm{ACC}(G,P,W;\Omega|_{G=c} \cap \Omega|_{i\in[r,r+19]}),$$

respectively.

Note that the introduced metrics can be computed for individual radar images and corresponding ground truth images, as well as for datasets $\mathcal{D} = \{(A_n, G_n)\}_{n=1}^N \subset D \times \{0,1\}^{n_\mathrm{r}\times n_\mathrm{a}}$ consisting of $N \in \mathbb{N}$ such image pairs. In this case, the metric is first computed for each radar image in the dataset individually and then aggregated as an area-weighted sum. For example, the area-weighted accuracy ACC on subsets $\Omega' = (\Omega'_n)_{n=1}^N \in \Omega^N$ for the dataset $\mathcal{D}$ is given by

$$\mathrm{ACC}(\mathcal{D};\Omega') = \frac{1}{\sum_{n=1}^N \mathcal{A}(\Omega'_n, W_n)} \sum_{n=1}^N \mathcal{A}(\Omega'_n, W_n)\,\mathrm{ACC}(G_n, P_n, W_n; \Omega'_n),$$

where $P_n \in \{0,1\}^{n_\mathrm{r}\times n_\mathrm{a}}$ is the prediction and $W_n \in \mathbb{R}^{n_\mathrm{r}\times n_\mathrm{a}}$ is the weight matrix for the $n$-th radar image. As before, setting $\Omega'_n = \Omega$ for all $n \in \{1,\dots,N\}$ yields the area-weighted accuracy for the dataset $\mathcal{D}$, while the evaluation on $\Omega'_n = \Omega|_{P_n=c}$ and $\Omega'_n = \Omega|_{G_n=c}$ for all $n \in \{1,\dots,N\}$ and $c \in \{0,1\}$ yields the class-specific precision and recall for the dataset $\mathcal{D}$, respectively. Analogously, the intersection over union IOU for the dataset $\mathcal{D}$ is given by

$$\mathrm{IOU}(\mathcal{D}) = \frac{1}{\sum_{n=1}^N \mathcal{A}(\Omega, W_n)} \sum_{n=1}^N \mathcal{A}(\Omega, W_n)\,\mathrm{IOU}(G_n, P_n, W_n).$$

# 3 State-of-the-art method

In this section, we describe the SotA method of DWD, which will be used as the benchmark for the evaluation in Sect. 4. It comprises a series of processing steps designed to identify and filter non-meteorological echoes, as well as a correction for attenuation, as described in Helmert et al. (2014). These processing steps include a threshold filter, a filter for spokes and rings, a blacklist filter, a polarimetric clutter filter based on fuzzy logic, a second-trip filter, and a speckle filter. Notably, the polarimetric clutter filter was extended in 2025 to include an additional so-called insect filter, which is used operationally during the summer months (April – October).

Echoes from insects, which are referred to as insect echoes, are typically characterized by the following features in the radar moments: reduced co-polar cross-correlation coefficient $\rho_\mathrm{HV}$ ($< 0.9$), increased differential reflectivity ZDR ($> 2\,\mathrm{dB}$), and low reflectivity factors DBZH and DBZV ($< 20\,\mathrm{dBZ}$). In principle, most pixels corresponding to insect echoes can be correctly classified based on these features. However, there are also weather events, such as snowfall, that exhibit similar features, particularly at greater distances from the radar. Moreover, pixels corresponding to a mixture of meteorological echoes and clutter can also exhibit these features but should be classified as meteorological echoes.

The goal of the DWD insect filter is to filter out insect echoes, i.e., identify the corresponding pixels and classify them as non-meteorological echoes. It utilizes the radar moments introduced previously, as well as the so-called clutter correction moment (CCORH), which quantifies the portion of the reflectivity filtered out by the Doppler filter in dBZ. The insect filter is a decision tree-based pixel-wise classifier that considers radar moments and the statistical properties of the local neighborhood.

More precisely, the newly implemented insect filter operates in two steps: First, all pixels eligible for filtering are identified. Then, for each pixel identified in the first step, it is checked whether the pixel will be filtered based on UDR. In both steps, the so-called $q \times q$ neighborhood $N_{(i,j)}^{q\times q} \subset \Omega$ of pixel $(i,j) \in \Omega$ for an odd integer $q \in \{1,3,5,\ldots\}$ is used, which is given by

$$N_{(i,j)}^{q\times q} = \left\{ (k,\ell) \in \Omega : |i-k| \leq \frac{q-1}{2} \text{ and } d(j,\ell) \leq \frac{q-1}{2} \right\},$$

where $d$ denotes the periodic distance measure as defined in Eq. (2). To incorporate information from the surroundings of a pixel into the filtering process, it has proven helpful to use percentile values. In contrast to using means, maxima, or minima, percentiles are insensitive to individual extreme values.

A pixel $(i,j) \in \Omega$ is eligible for filtering if (a) its reflectivity value is below 20 dBZ, (b) the CCORH value in its $3\times3$ neighborhood $N_{(i,j)}^{3\times3}$ is not significant, and (c) the ZDR value in $N_{(i,j)}^{3\times3}$ is sufficiently different from 0 dB. All three conditions must be met simultaneously to prevent the misclassification of meteorological echoes. In particular, the CCORH condition was introduced to retain pixels within precipitation areas that, due to a mixture of meteorological echoes and clutter, exhibit a reduced quality of the polarimetric moments. Otherwise, they would be filtered out based on their high UDR value. Furthermore, the considered radar moments of snow and insects are very similar, except for ZDR. Therefore, the ZDR condition is necessary to retain pixels corresponding to weak snowfall, especially at greater distances from the radar.

To avoid filtering meteorological echoes at greater distances, where the reduced signal-to-noise ratio results in lower-quality polarimetric moments, an additional distance-dependent condition is implemented. Pixels at greater distances from the radar are only eligible for filtering if a certain percentage of pixels closer to the radar have already been filtered out by the insect filter. This condition accounts for the fact that insects usually occur in near-surface layers, and if insects are present at higher altitudes, they are typically also found closer to the ground.

An eligible pixel is filtered out if either the pixel itself has a high UDR value ($> -5$ dB), or if high UDR values dominate in the local neighborhood (20th-percentile UDR in its $5 \times 5$ neighborhood $> -11$ dB) and the pixel itself does not have a particularly low value ($< -18$ dB). These two conditions ensure that pixels with a very high likelihood of non-meteorological echoes (based on their UDR values), as well as those mainly surrounded by pixels corresponding to non-meteorological echoes with moderate UDR values, are filtered out.

Note that the distance-dependent condition and the thresholds for conditions (b) and (c) might have to be chosen individually for each radar station, as they depend on the radar wavelength, the scan strategy and the surrounding topography, which determine the exact characteristics of non-meteorological echoes. We therefore do not specify any explicit values in this paper.

# 4 Results

In this section, we first present the outcome of the model selection described in Sect. 2.6 and then evaluate the performance of the selected network $f_{\theta^*}$ through both visual inspection and quantitative analysis of the predicted segmentations. Although all mixed radar images used for training were synthetically generated, we evaluate network performance on a test dataset of experimentally measured mixed radar images with expert-labeled ground truth. Furthermore, we benchmark these results against those achieved by the SotA method described in Sect. 3 and investigate the importance of the individual radar moments in an ablation study.

Figure 8 shows the validation losses $L^{\mathrm{val}}_{n_\mathrm{f},e}$ and their moving averages $\bar{L}^{\mathrm{val}}_{n_\mathrm{f},e}$ during training for the four networks with $n_\mathrm{f} \in \{8, 16, 32, 64\}$ feature channels. The smallest moving average of the validation loss is attained by the network with $n_\mathrm{f} = 64$ feature channels after 90 epochs, i.e., with $n_\mathrm{f}^* = 64$ and $e^* = 90$. Its trainable parameters $\theta^* = \theta_{64,90}$ are used in the following evaluation.

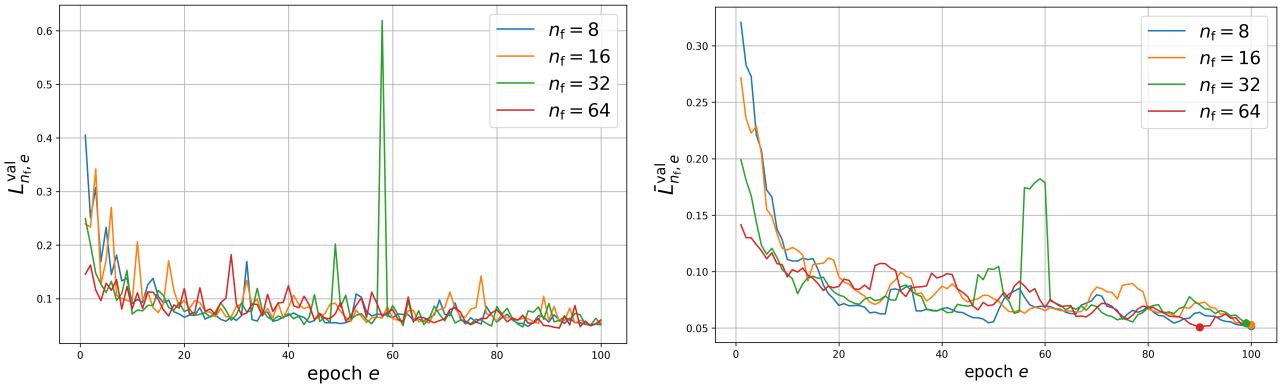


Figure 8: Validation loss $L^{\mathrm{val}}_{n_\mathrm{f},e}$ (left) and its moving average $\bar{L}^{\mathrm{val}}_{n_\mathrm{f},e}$ (right) over the epochs $e$ for each number of feature channels $n_\mathrm{f} \in \{8, 16, 32, 64\}$.

We evaluate the trained network using the metrics stated in Sect. 2.7 and compare its performance to the SotA method detailed in Sect. 3. Segmentation performance is quantified on training datasets $\mathcal{D}^{\mathrm{m}}_{\mathrm{train}}$, $\mathcal{D}^{\mathrm{w}}_{\mathrm{train}}$, and $\mathcal{D}^{\mathrm{s}}_{\mathrm{train}}$, and validation datasets $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$, $\mathcal{D}^{\mathrm{w}}_{\mathrm{val}}$, and $\mathcal{D}^{\mathrm{s}}_{\mathrm{val}}$, as introduced in Sect. 2.5. Moreover, both methods are also evaluated on a test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$ of experimentally measured mixed radar images, given by

$$\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}} = \big\{(A_n, G_n) : n \in \{1, \ldots, 5\}\big\},$$

where $A_n \in D$ are radar images measured in June 2022 featuring both meteorological and non-meteorological echoes, and $G_n \in \{0, 1\}^{n_\mathrm{r} \times n_\mathrm{a}}$ are the corresponding ground truth images, which were manually labeled by a subject expert. Two of the five images were measured at 09:00 UTC on June 5, 2022 (radar stations Essen and Neuhaus), one at 09:00 UTC on June 16, 2022 (Dresden), and two at 16:00 UTC on June 30, 2022 (Neuheilenbach and Türkheim). The five test images are shown in Fig. 9 and Fig. 1b. They were selected to cover a variety of scenarios: stratiform and convective precipitation, convection close to the radar and at longer ranges, morning and afternoon data, and different radars. The test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$ was not used for training or model selection. However, the test image from June 16 was measured during the one-hour period from which the reflectivity weighting factor in Sect. 2.4 is determined.

Table 2 shows a comparison of the resulting metrics averaged over the respective datasets as described in Sect. 2.7. The overall performance metrics on the validation datasets are slightly lower than those on the training datasets. The accuracy ACC and intersection over union IOU on the mixed validation dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$ are 0.972 and 0.946, respectively, compared to 0.989 and 0.979 on the mixed training dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{train}}$.

Figure 10 visualizes the echoes that were correctly and incorrectly classified by the network for one mixed validation image in dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$. Most of the meteorological echoes, particularly those with high reflectivity, were correctly identified. This aligns with the recall value $\mathrm{REC}^0$ for meteorological echoes of 0.972 on dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$. Moreover, the resulting segmentation does not show any notable holes within areas surrounded by meteorological echoes. However, as can

be seen in Fig. 10c, misclassified non-meteorological echoes are clearly visible as small speckles, leading to a lower precision value $\mathrm{PRE}^0$ for meteorological echoes of 0.936.

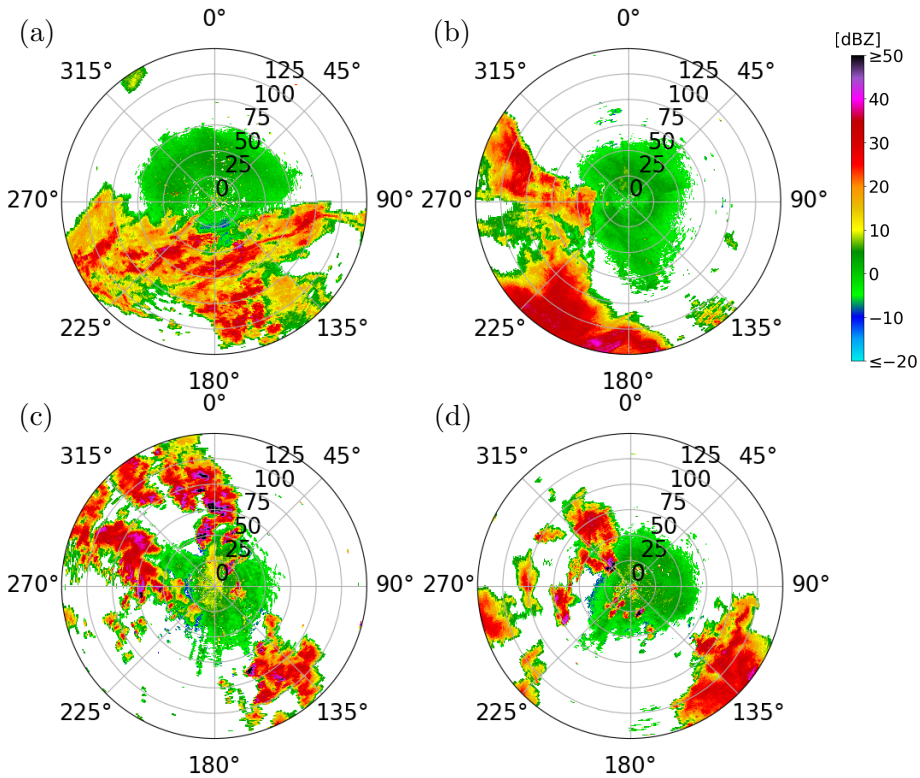


Figure 9: Overview of four of the five DBZH radar images in the test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$. The radar images shown were measured (a) on June 5, 2022, at 09:00 UTC in Essen, (b) on June 5, 2022, at 09:00 UTC in Neuhaus, (c) on June 30, 2022, at 16:00 UTC in Neuheilenbach and (d) on June 30, 2022, at 16:00 UTC in Türkheim. The remaining radar image from $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$, which was measured on June 16, 2022, at 09:00 UTC in Dresden, is shown in Fig. 1b.

Table 2: Area-weighted performance metrics for the evaluation of the trained network and the SotA method. Precision $\mathrm{PRE}^c$ and recall $\mathrm{REC}^c$ are computed for meteorological ($c = 0$) and non-meteorological echoes ($c = 1$). For the test dataset, the better value of the two methods is printed in bold.

| Method | Type of data | $\mathrm{PRE}^0$ | $\mathrm{PRE}^1$ | $\mathrm{REC}^0$ | $\mathrm{REC}^1$ | ACC | IOU |
|---|---|---|---|---|---|---|---|
| network | mixed synthetic images (training) | 0.993 | 0.976 | 0.992 | 0.979 | 0.989 | 0.979 |
| | cluttered summer images (training) | - | - | - | 0.992 | 0.992 | 0.992 |
| | winter images (training) | - | - | 0.997 | - | 0.997 | 0.997 |
| | mixed synthetic images (validation) | 0.936 | 0.988 | 0.972 | 0.972 | 0.972 | 0.946 |
| | cluttered summer images (validation) | - | - | - | 0.987 | 0.987 | 0.987 |
| | winter images (validation) | - | - | 0.994 | - | 0.994 | 0.994 |
| | mixed measured images (test) | 0.967 | **0.959** | **0.975** | 0.946 | **0.964** | **0.930** |
| SotA | mixed measured images | **0.990** | 0.789 | 0.837 | **0.986** | 0.894 | 0.811 |

When evaluated on the experimentally measured mixed dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$, the network performance is comparable to that on the validation dataset, with accuracy ACC and intersection over union IOU of 0.964 and 0.930, respectively. However, when evaluating the SotA method on the

same mixed dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$, it performs substantially worse, with accuracy ACC and intersection over union IOU of 0.894 and 0.811, respectively. These overall differences are caused by larger differences in precision $\mathrm{PRE}^1$ for non-meteorological echoes and recall $\mathrm{REC}^0$ for meteorological echoes, where the SotA method achieves scores of 0.789 and 0.837, respectively, compared to the neural network with scores of 0.959 and 0.975. The SotA method achieves better scores for recall $\mathrm{REC}^1$ for non-meteorological echoes and precision $\mathrm{PRE}^0$ for meteorological echoes, with values of 0.986 and 0.990, respectively, compared to the neural network with scores of 0.946 and 0.967. Table 3 shows the performance metrics for each of the five radar images of the test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$. The intersection over union IOU of the network ranges from 0.855 (Neuheilenbach) to 0.974 (Essen). For each labeled radar image, the intersection over union IOU and the accuracy ACC are higher than the respective values for the SotA method.

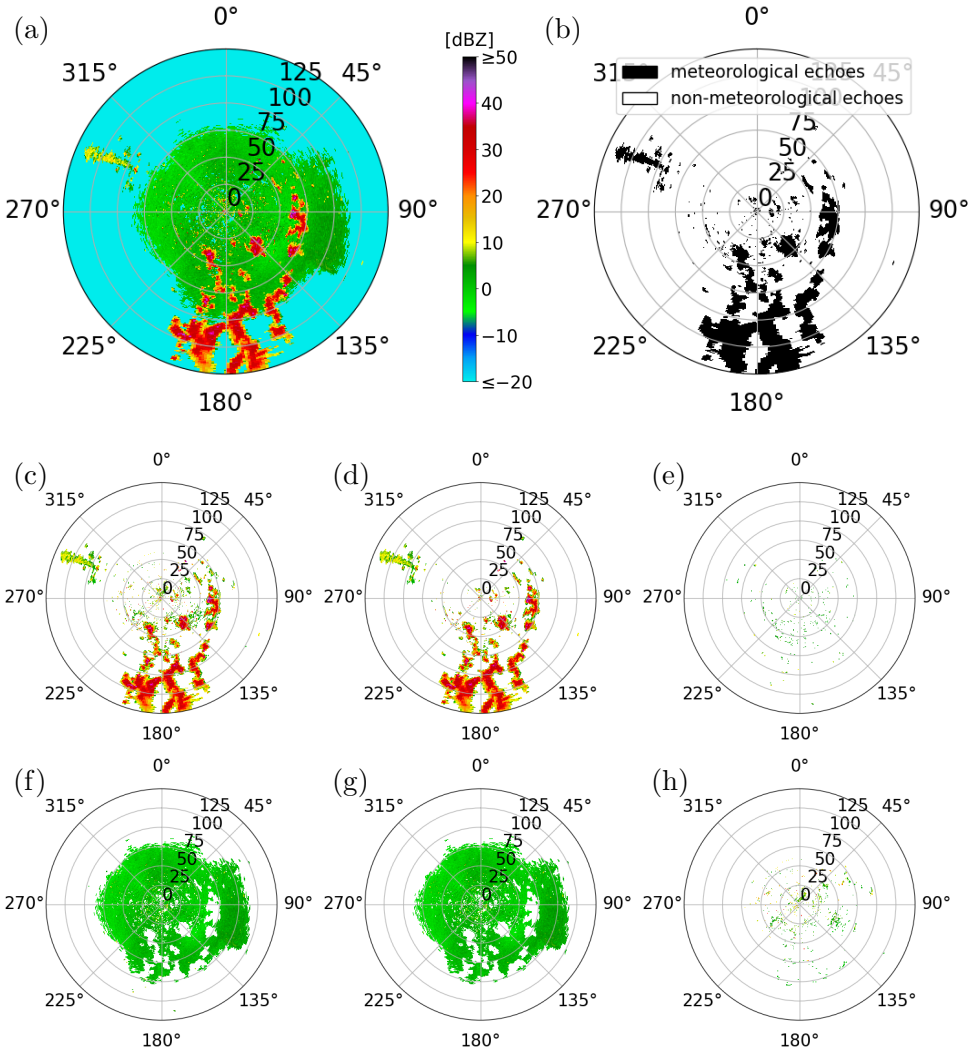


Figure 10: Evaluation of the network on a synthetic validation image: (a) input image and (b) corresponding ground truth, (c) pixels classified as meteorological echoes by the network, (d) correctly classified meteorological echoes, and (e) meteorological echoes misclassified as non-meteorological echoes. The bottom row shows (f) pixels classified as non-meteorological echoes, (g) correctly classified non-meteorological echoes, and (h) non-meteorological echoes misclassified as meteorological echoes. The radar image shown was generated using two radar images measured on January 9, 2025, at 17:05 UTC and on June 18, 2022, at 10:55 UTC in Essen.

Table 3: Evaluation of the trained network and the SotA method. The table shows the area-weighted performance metrics for each of the five radar images of the test dataset $\mathcal{D}_{\text{exp}}^{\text{m}}$. Precision $\text{PRE}^c$ and recall $\text{REC}^c$ are computed for meteorological ($c = 0$) and non-meteorological echoes ($c = 1$). For each radar image, the better value of the two methods is printed in bold.

| Method | Date and time | Radar station | $\text{PRE}^0$ | $\text{PRE}^1$ | $\text{REC}^0$ | $\text{REC}^1$ | ACC | IOU |
|---|---|---|---|---|---|---|---|---|
| network | 05.06.2022, 09:00 UTC | Essen | 0.986 | **0.990** | **0.996** | 0.965 | **0.987** | **0.974** |
| | 05.06.2022, 09:00 UTC | Neuhaus | 0.982 | **0.930** | **0.956** | 0.971 | **0.962** | **0.927** |
| | 16.06.2022, 09:00 UTC | Dresden | 0.969 | **0.958** | **0.926** | 0.983 | **0.962** | **0.927** |
| | 30.06.2022, 16:00 UTC | Neuheilenbach | 0.917 | **0.948** | **0.986** | 0.744 | **0.923** | **0.855** |
| | 30.06.2022, 16:00 UTC | Türkheim | **0.988** | **0.969** | **0.984** | **0.976** | **0.982** | **0.964** |
| SotA | 05.06.2022, 09:00 UTC | Essen | **0.997** | 0.787 | 0.888 | **0.993** | 0.919 | 0.855 |
| | 05.06.2022, 09:00 UTC | Neuhaus | **0.991** | 0.787 | 0.840 | **0.987** | 0.895 | 0.813 |
| | 16.06.2022, 09:00 UTC | Dresden | **0.984** | 0.841 | 0.678 | **0.994** | 0.878 | 0.776 |
| | 30.06.2022, 16:00 UTC | Neuheilenbach | **0.993** | 0.702 | 0.853 | **0.982** | 0.887 | 0.808 |
| | 30.06.2022, 16:00 UTC | Türkheim | 0.977 | 0.758 | 0.848 | 0.960 | 0.885 | 0.800 |

Figure 11 shows the incorrectly segmented echoes of the network (Figs. 11c and 11f) and of the SotA method (Figs. 11b and 11e) for one experimentally measured mixed radar image in dataset $\mathcal{D}_{\text{exp}}^{\text{m}}$. The top row shows the incorrect segmentations of meteorological echoes for both methods, with the neural network making substantially fewer errors. Here, the SotA method tends to incorrectly classify the pixels located at the boundary of weather events. However, the SotA method performs substantially better in correctly identifying non-meteorological echoes near the radar station.

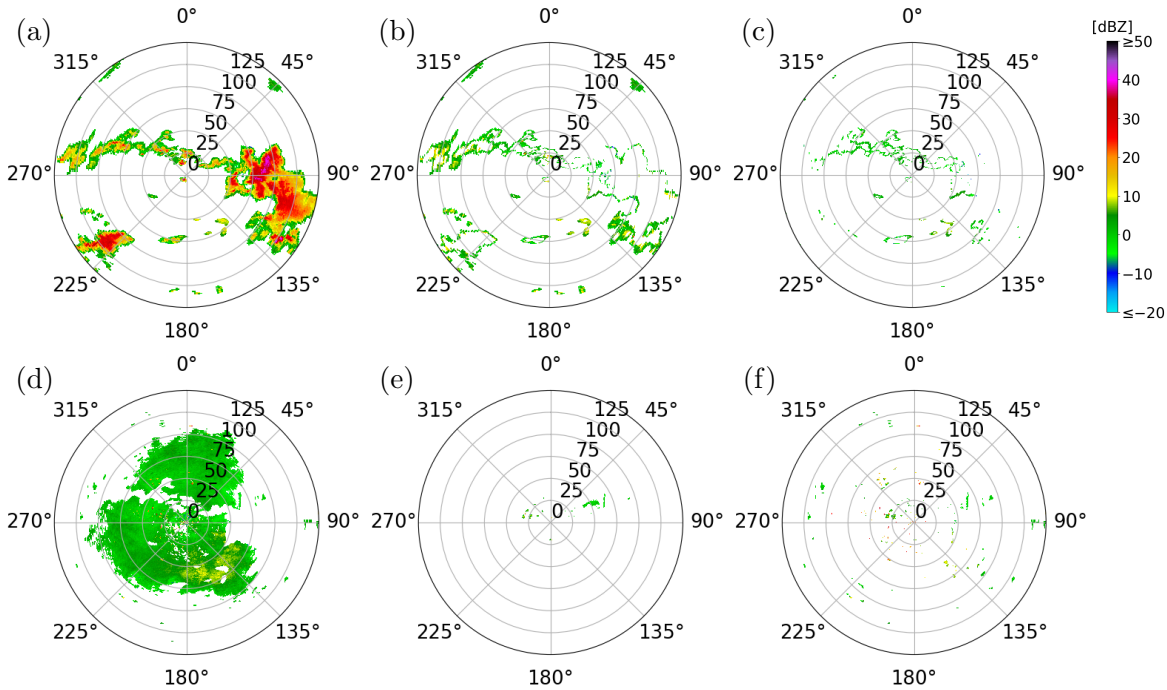


Figure 11: Evaluation of the SotA method and the trained network on the experimentally measured radar image shown in Fig. 1b: The first row shows (a) the ground truth for the meteorological echoes and the meteorological echoes misclassified as non-meteorological echoes by (b) the SotA method and (c) the trained network. Analogously, the second row shows (d) the hand-labeled ground truth for the non-meteorological echoes and the non-meteorological echoes misclassified as meteorological echoes by (e) the SotA method and (f) the trained network.

In order to further investigate this, Fig. 12a shows the accuracy depending on the distance, while Fig. 12b and c show the class-specific recall scores for meteorological and non-meteorological echoes, respectively. The network's accuracy $\mathrm{ACC}^{\mathrm{D}}$ on the test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$ increases with distance from the radar, rising from below 0.9 within 25 km to nearly 1.0 beyond 75 km. The same trend can be observed for the validation dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$, for which only slightly higher scores are achieved by the network. When compared to the SotA method, the recall $\mathrm{REC}^{0,\mathrm{D}}$ for meteorological echoes depending on distance is higher for the network than for the SotA method for all distances, and the opposite is true for non-meteorological echoes. The most substantial differences occur close to the radar, at a distance of less than 25 km, where the neural network achieves a recall $\mathrm{REC}^{1,\mathrm{D}}$ of 0.885, compared to that of the SotA method with 0.981. This relatively lower recall close to the radar quantifies the errors we see in Fig. 11f, where a substantial number of non-meteorological echoes near the radar are incorrectly classified. When considering the recall $\mathrm{REC}^{0,\mathrm{D}}$ for meteorological echoes depending on distance, the network clearly outperforms the SotA method at distances further from the radar, with a recall value of 0.986 compared to that of the SotA method with 0.831 beyond 75 km. The echoes further away from the radar represent a proportionally larger area, as shown in the histograms, and thus the network outperforms the SotA method overall.

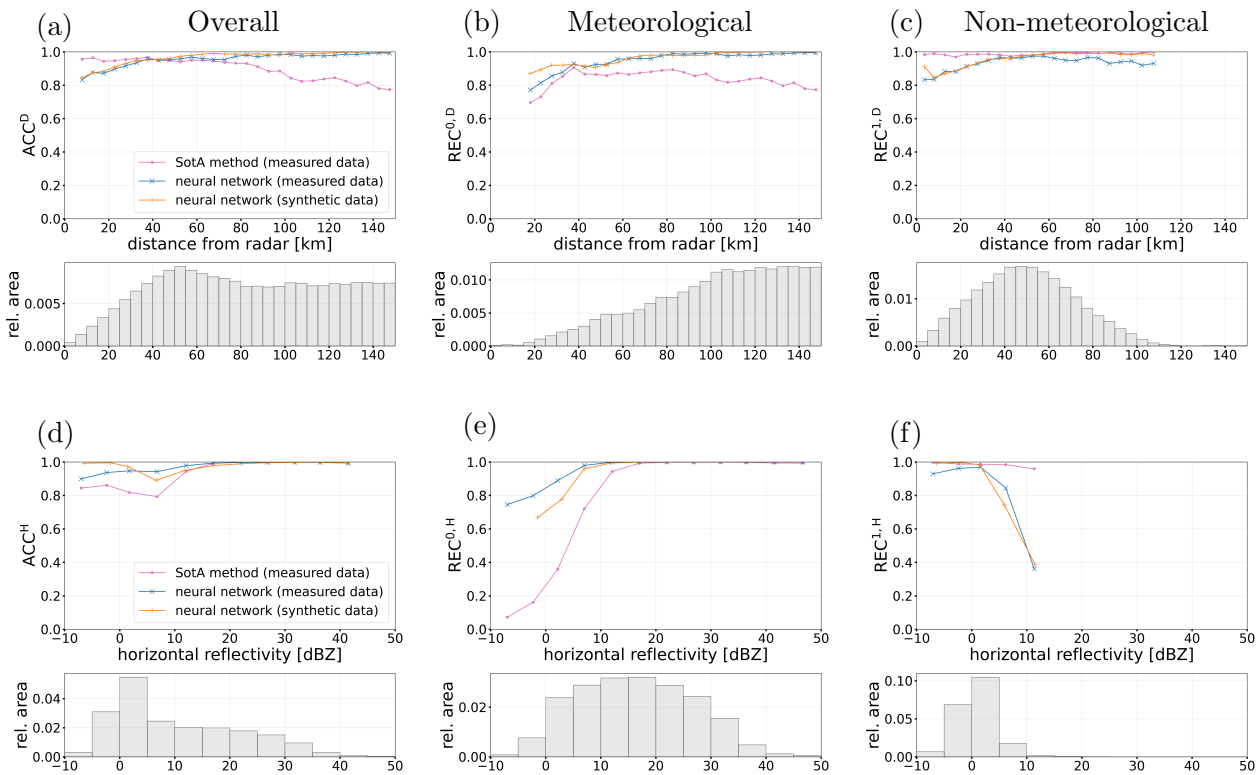


Figure 12: First row: Accuracy in dependence on the distance from the radar for (a) meteorological and non-meteorological echoes combined, (b) meteorological echoes, and (c) non-meteorological echoes, as classified in the ground truth. Second row: Accuracy in dependence on the horizontal reflectivity values for the same three sets of pixels. The metrics are only shown for distances and horizontal reflectivities with a relative area greater than 0.0005, respectively. The histograms below show the area distribution for each of the three sets of pixels, respectively. The SotA method was evaluated on the experimentally measured mixed dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$ and the neural network was evaluated on the same experimentally measured dataset and additionally on the mixed validation dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$.

Analogously, we evaluate the accuracy and class-specific recall in dependence on the horizontal reflectivity values, as shown in Fig. 12d to f. For meteorological echoes with a reflectivity higher than 10 dBZ, both the network and the SotA method achieve a recall $\mathrm{REC}^{0,\mathrm{H}}$ near 1.0. However, for horizontal reflectivities below 10 dBZ, the SotA performance degrades substantially, whereas the network maintains a recall of 0.925. While meteorological echoes with reflectivity values below 0 dBZ are rare in the dataset, those with reflectivity values between 0 and 10 dBZ are much more frequent, as shown in the histogram of Fig. 12e. For example, the edges of weather events are typically less reflective and are often incorrectly classified as non-meteorological echoes by the SotA method, as shown in Fig. 11b. In the case of non-meteorological echoes, the network struggles to classify echoes with high reflectivity ($> 10$ dBZ), achieving a lower recall $\mathrm{REC}^{1,\mathrm{H}}$ of 0.308 compared to the SotA method, which achieves a recall value of 0.961. Nevertheless, such high-reflectivity non-meteorological echoes are rare in the dataset.

To investigate the influence of the individual input channels, one network is trained for each of the seven combinations of the DBZH, ZDR, and UDR channels. All networks are trained on the same synthetic training data and with the same hyperparameters as described in Sect. 2.5, where the number of feature channels is fixed to $n_\mathrm{f} = 64$ and the number of training epochs is selected based on the validation loss as described in Sect. 2.6.

Table 4 shows the resulting performance metrics on the mixed validation dataset $\mathcal{D}^\mathrm{m}_\mathrm{val}$ and the test dataset $\mathcal{D}^\mathrm{m}_\mathrm{exp}$, where the row using all three channels corresponds to the network considered previously. Among the networks using a single input channel, the network using the DBZH channel achieves the highest metrics on the test dataset. The highest metrics are achieved by the network combining the DBZH and UDR channels, with an accuracy of 0.966 and an intersection over union of 0.933 on the test dataset. When DBZH and UDR channels are considered, the ZDR channel does not appear to improve performance on the test dataset. Since UDR is deduced from ZDR and $\rho_\mathrm{HV}$, this result suggests that the $\rho_\mathrm{HV}$ contributions to UDR add to the performance, whereas using both UDR and ZDR does not further improve the performance.

Table 4: Area-weighted accuracy ACC and intersection over union IOU for the trained networks in dependence on the input channels used. The networks are evaluated on the mixed validation dataset $\mathcal{D}^\mathrm{m}_\mathrm{val}$ and the test dataset $\mathcal{D}^\mathrm{m}_\mathrm{exp}$. The best value in each column is printed in bold.

| Used input channels | | | Metrics ($\mathcal{D}^\mathrm{m}_\mathrm{val}$) | | Metrics ($\mathcal{D}^\mathrm{m}_\mathrm{exp}$) | |
|---|---|---|---|---|---|---|
| DBZH | ZDR | UDR | ACC | IOU | ACC | IOU |
| ✓ | | | 0.953 | 0.911 | 0.961 | 0.925 |
| | ✓ | | 0.966 | 0.935 | 0.940 | 0.885 |
| | | ✓ | 0.934 | 0.879 | 0.911 | 0.834 |
| ✓ | ✓ | | 0.963 | 0.930 | 0.955 | 0.913 |
| ✓ | | ✓ | 0.958 | 0.921 | **0.966** | **0.933** |
| | ✓ | ✓ | 0.971 | 0.944 | 0.949 | 0.903 |
| ✓ | ✓ | ✓ | **0.972** | **0.946** | 0.964 | 0.930 |

# 5 Discussion

The results presented in Sect. 4 demonstrate that a neural network is capable of effectively identifying non-meteorological echoes in radar images. One of the most significant findings is the network's ability to generalize from synthetically generated radar images to experimentally measured mixed radar images $\mathcal{D}^\mathrm{m}_\mathrm{exp}$. Although the training process relied exclusively on synthet-

ically generated radar images, network performance on experimentally measured radar images was in line with, or exceeding, that of the SotA method currently used at DWD.

A key advantage of the proposed network-based method over the SotA method used at DWD lies in the preservation of meteorological echoes, particularly at the boundaries of weather events and in regions of weak precipitation. As illustrated in Fig. 11b, and quantified in Fig. 12e, the SotA method tends to incorrectly classify the boundaries of weather events as non-meteorological echoes. This is likely due to the SotA method's reliance on heuristically chosen thresholds and local neighborhood statistics (e.g., $3 \times 3$ or $5 \times 5$ neighborhoods). In contrast, the network utilizes context from a larger neighborhood, allowing it to better recognize the spatial structures of weather events.

However, the evaluation also highlighted limitations that should be addressed in future work. The network showed a lower recall $\mathrm{REC}^1$ for non-meteorological echoes compared to the SotA method, particularly for radar gates close to the radar at distances of less than 25 km, as shown in Fig. 12c. The network's errors are likely a consequence of the underlying winter radar images used for training. For the generation of mixed synthetic radar images, we assign echoes in winter radar images exclusively as meteorological echoes. However, in reality, the winter radar images also contain non-meteorological echoes, particularly in areas close to the radar, as can be seen in Fig. 3. Consequently, the corresponding pixels are incorrectly labeled as meteorological in the ground truth. Therefore, the network may be disincentivized from correctly identifying non-meteorological echoes near the radar during training. A quantitative verification of this explanation, as well as possible improvements of the network's performance near the radar, is left for future work.

The contributions of the individual input channels are investigated by the ablation study in Sect. 4, see Table 4. As expected, the horizontal reflectivity appears the most informative radar moment, with an intersection over union of 0.925 on the test dataset, compared to 0.885 for ZDR and 0.834 for UDR. For the depolarization ratio, it was unclear whether the synthetic UDR images are useful for training at all, since each of their values is taken from only one of the two combined radar images, see Sect. 2.4. However, appending the UDR channel to the DBZH channel increases the intersection over union on the test dataset to 0.933. Moreover, the network trained using this combination performs similarly on the validation dataset, with an intersection over union of 0.921. This suggests that what the network learns from the synthetic UDR images transfers to experimentally measured radar images. In addition, this result suggests that a physically interpretable combination method is not necessarily required for an input channel to be useful. Additional radar moments could therefore be included in the training data by the same simple assignment used for the synthetic UDR images, even if a physically motivated mixing of two radar images is not available. In particular, Doppler-based moments, such as the radial velocity or the clutter correction moment used by the SotA method, are promising candidates.

Regarding the model's generalizability, the evaluation results indicate that the network did not suffer from substantial overfitting during training. The overall performance metrics on the mixed training dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{train}}$ are similar to those observed on the mixed validation dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{val}}$ and the test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$. The mixed training data $\mathcal{D}^{\mathrm{m}}_{\mathrm{train}}$ was generated by combining cluttered summer radar images measured on June 18, 2022 (10:00–11:00 UTC) with winter radar images measured on January 9, 2025 (17:00–18:00 UTC). The weighting factors $\lambda_{\mathrm{w}}$ and $\lambda_{\mathrm{s}}$ used to combine these were explicitly calibrated so that the generated images resembled mixed summer radar images measured on June 16, 2022 (09:00–10:00 UTC). The validation datasets were generated from radar images of the same time periods, measured at a radar station that was not used for

training, see Sect. 2.5. In contrast, the test dataset $\mathcal{D}^{\mathrm{m}}_{\mathrm{exp}}$ consists of five experimentally measured mixed radar images from three different days. Four of them were measured on June 5 and June 30, 2022, i.e., on days from which no radar data was used for the generation of the synthetic training data, and the two images from June 30 were measured in the afternoon (16:00 UTC). The remaining test image was measured on June 16, 2022, during the one-hour period from which the weighting factors were determined. The test dataset contains a variety of stratiform and convective precipitation, close to the radar and at longer ranges. The network performed well on all five test images, see Table 3. This indicates that the network transfers to experimentally measured summer images of other days and times of day, although robustness across seasons and event types cannot be established with five images from three summer days. Moreover, all meteorological echoes in the training data are taken from the winter radar images. Artifacts, such as differential attenuation, that occur in strong summertime convection are therefore absent from the training data. The five test images are unlikely to represent the wide range of such rare artifacts, and evaluating the performance of the network on them requires a larger amount of test data. To further enhance and evaluate the network's generalization capabilities, future work should incorporate data from a broader range of meteorological situations. Despite the very limited amount of data for training, the network already shows good performance. For operational purposes, the training dataset will have to be extended substantially, such that it incorporates all kinds of possible meteorological situations and artifacts. For example, summer-only artifacts could be incorporated by partially labeling a small number of radar images in which they occur, or by generating them synthetically.

There are several improvements to the proposed method. First, as seen in Fig. 10, the network output includes isolated misclassified pixels. Applying a post-processing despeckle filter, similar to the one described in Sect. 2.2 for the input data, could remove these small artifacts and further improve precision without adding methodological complexity. Second, the current approach processes radar images from a single elevation angle and time step. Incorporating additional elevation angles or time steps could provide the network with additional spatially-resolved or time-resolved structural information. Although this would add significant complexity to the proposed methodology, it represents a very promising adjustment. For one, non-meteorological echoes are less common at higher elevation angles, whereas weather events often extend vertically. Furthermore, incorporating time-resolved information would allow the network to exploit differences in movement patterns. For example, insects are typically either stationary or move irregularly, whereas weather events exhibit coherent motion (Fig. 13).

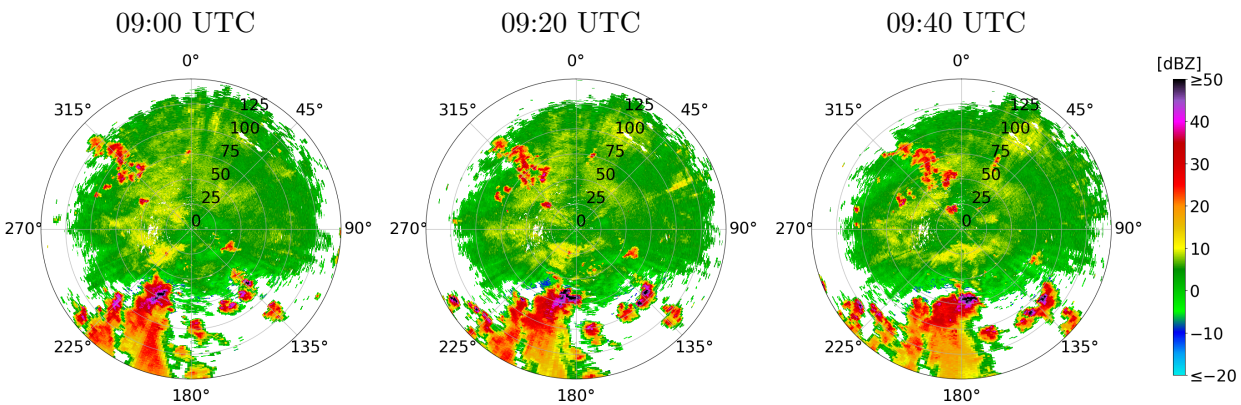


Figure 13: Time series of mixed summer radar images collected at the same radar station at 20-minute intervals. The radar images shown were measured on June 16, 2022 in Isen.

# 6 Conclusions

Accurate identification of non-meteorological echoes in radar images is a prerequisite for reliable quantitative precipitation estimation, hydrometeor classification and aviation safety. In this work, we introduced a deep learning approach using a U-net architecture to segment dual-polarization weather radar data. To overcome the scarcity of pixel-wise labeled training data, we developed a novel method for generating synthetic training data by combining winter data and cluttered summer data.

The proposed method was evaluated against a state-of-the-art (SotA) method currently used at Deutscher Wetterdienst. Our results show that the neural network outperforms the SotA method in preserving weak precipitation and the boundaries of weather events, achieving an overall accuracy of 0.964 and an intersection over union of 0.930, compared to 0.894 and 0.811 for the SotA method on five expert-labeled radar images.

Most importantly, this study demonstrates that synthetically generated radar moments can be used for training deep neural networks in order to accurately identify non-meteorological echoes. Despite being trained exclusively on synthetically generated mixed radar data, the network accurately identified non-meteorological echoes in experimentally measured mixed radar data without any further fine-tuning.

To ensure robustness across the entire year, future work should incorporate data from spring and autumn to capture seasonal variations. Furthermore, the neural network's capabilities could be extended beyond 2D spatial features. Utilizing consecutive time steps or multiple elevation angles as input could provide the time-resolved and volumetric context needed for even more precise segmentation.

# Code and data availability

The code and data used for the current study are available from the corresponding authors on reasonable request.

# Author contribution

Conceptualization: AB, RB, LF, TK. Data curation: AB, RB. Formal analysis: RB, LF, TK. Investigation: AB, RB. Methodology: RB, LF, TK. Project administration: UB, VS. Resources: UB, AB, VS. Software: AB, RB, TK. Supervision: UB, VS. Validation: RB, TK. Visualization: RB, TK. Writing (original draft preparation): UB, AB, RB, LF, TK. Writing (review and editing): all authors.

# Acknowledgments

The authors thank the reviewers (Matteo Guidicelli and Jenna Ritvanen) for their constructive comments and suggestions, which significantly improved the quality of this manuscript.

# References

Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., Chauhan, G., Chourdia, A., Constable, W., Desmaison, A., DeVito, Z., Elli-

son, E., Feng, W., Gong, J., Gschwind, M., Hirsh, B., Huang, S., Kalambarkar, K., Kirsch, L., Lazos, M., Lezcano, M., Liang, Y., Liang, J., Lu, Y., Luk, C. K., Maher, B., Pan, Y., Puhrsch, C., Reso, M., Saroufim, M., Siraichi, M. Y., Suk, H., Suo, M., Tillet, P., Wang, E., Wang, X., Wen, W., Zhang, S., Zhao, X., Zhou, K., Zou, R., Mathews, A., Chanan, G., Wu, P., and Chintala, S. (2024). PyTorch 2: Faster machine learning through dynamic Python bytecode transformation and graph compilation. In Gupta, R., Abu-Ghazaleh, N. B., Musuvathi, M., and Tsafrir, D., editors, *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS 2024)*, pages 929–947. ACM.

Atanbori, J., Frantzidis, C. A., Al-Khafajiy, M., Aliyu, A., Sohani, B., Appiah, K., Moore, H., Sanders, C., and Ward, A. I. (2025). Learning with noisy labels for classifying biological echoes in polarimetric weather radar observations using artificial neural networks. *Neurocomputing*, 634:129892.

Bauer, S., Tielens, E. K., and Haest, B. (2024). Monitoring aerial insect biodiversity: A radar perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 379:20230113.

Besic, N., Figueras i Ventura, J., Grazioli, J., Gabella, M., Germann, U., and Berne, A. (2016). Hydrometeor classification through statistical clustering of polarimetric radar measurements: A semi-supervised approach. *Atmospheric Measurement Techniques*, 9:4425–4445.

Bringi, V. N. and Chandrasekar, V. (2001). *Polarimetric Doppler Weather Radar: Principles and Applications*. Cambridge University Press.

Chen, J.-Y., Trömel, S., Ryzhkov, A., and Simmer, C. (2021). Assessing the benefits of specific attenuation for quantitative precipitation estimation with a C-band radar network. *Journal of Hydrometeorology*, 22:2617–2631.

Doviak, R. J. and Zrnić, D. S. (1993). *Doppler Radar and Weather Observations*. Academic Press, 2nd edition.

Dufton, D. R. L. and Collier, C. G. (2015). Fuzzy logic filtering of radar reflectivity to remove non-meteorological echoes using dual polarization radar moments. *Atmospheric Measurement Techniques*, 8:3985–4000.

Dumitrescu, D. and Boiangiu, C.-A. (2019). A study of image upsampling and downsampling filters. *Computers*, 8:30.

Fabry, F. (2015). *Radar Meteorology: Principles and Practice*. Cambridge University Press.

Frech, M., Hagen, M., and Mammen, T. (2017). Monitoring the absolute calibration of a polarimetric weather radar. *Journal of Atmospheric and Oceanic Technology*, 34(3):599–615.

Gauthreaux, S. and Diehl, R. (2020). Discrimination of biological scatterers in polarimetric weather radar data: Opportunities and challenges. *Remote Sensing*, 12:545.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.

Helmert, K., Tracksdorf, P., Steinert, J., Werner, M., Frech, M., Rathmann, N., Hengstebeck, T., Mott, M., Schumann, S., and Mammen, T. (2014). DWDs new radar network and post-processing algorithm chain. In *8th European Conference on Radar in Meteorology*

*and Hydrology*, volume 4, pages 1–5. `https://www.pa.op.dlr.de/erad2014/programme/ExtendedAbstracts/237_Helmert.pdf`.

Hoshen, J. and Kopelman, R. (1976). Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. *Physical Review B*, 14:3438–3445.

Hu, C., Sun, Z., Cui, K., Mao, H., Wang, R., Kou, X., Wu, D., and Xia, F. (2024). Classification of biological scatters using polarimetric weather radar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:7436–7447.

Jatau, P., Melnikov, V., and Yu, T.-Y. (2021). A machine learning approach for classifying bird and insect radar echoes with S-band polarimetric weather radar. *Journal of Atmospheric and Oceanic Technology*, 38:1797–1812.

Kern, N., Aguilar, J., Grebner, T., Meinecke, B., and Waldschmidt, C. (2022). Learning on multistatic simulation data for radar-based automotive gesture recognition. *IEEE Transactions on Microwave Theory and Techniques*, 70:5039–5050.

Kilambi, A., Fabry, F., and Meunier, V. (2018). A simple and effective method for separating meteorological from nonmeteorological targets using dual-polarization data. *Journal of Atmospheric and Oceanic Technology*, 35:1415–1424.

Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization. arXiv:1412.6980.

Krauss, D., Engel, L., Ott, T., Braunig, J., Richer, R., Gambietz, M., Albrecht, N., Hille, E., Ullmann, I., Braun, M., Dabrock, P., Kolpin, A., Koelewijn, A., Eskofier, B., and Vossiek, M. (2024). A review and tutorial on machine learning-enabled radar-based biomedical monitoring. *IEEE Open Journal of Engineering in Medicine and Biology*, 6:680–699.

Leinonen, J., Hamann, U., Nerini, D., Germann, U., and Franch, G. (2023). Latent diffusion models for generative precipitation nowcasting with accurate uncertainty quantification. arXiv:2304.12891.

Mäkinen, T., Ritvanen, J., Pulkkinen, S., Weisshaupt, N., and Koistinen, J. (2022). Bayesian classification of nonmeteorological targets in polarimetric doppler radar measurements. *Journal of Atmospheric and Oceanic Technology*, 39:1561–1578.

Mbizvo, G. K. and Larner, A. J. (2024). On the dependence of the critical success index (CSI) on prevalence. *Diagnostics*, 14(5).

Melnikov, V. and Matrosov, S. Y. (2013). Estimations of aspect ratios of ice cloud particles with the WSR-88D radar. In *36th AMS Conference on Radar Meteorology*, pages 1–10.

Michelson, D., Hansen, B., Jacques, D., Lemay, F., and Rodriguez, P. (2020). Monitoring the impacts of weather radar data quality control for quantitative application at the continental scale. *Meteorological Applications*, 27:e1929.

Pulkkinen, S., Chandrasekar, V., von Lerber, A., and Harri, A.-M. (2020). Nowcasting of convective rainfall using volumetric radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):7845–7859.

Radhakrishna, B., Fabry, F., and Kilambi, A. (2019). Fuzzy logic algorithms to identify birds, precipitation, and ground clutter in S-band radar data using polarimetric and nonpolarimetric variables. *Journal of Atmospheric and Oceanic Technology*, 36:2401–2414.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597.

Ryzhkov, A. V. and Zrnić, D. S. (2019). *Radar Polarimetry for Weather Observations*. Springer.

Sauvageot, H. (1992). *Radar Meteorology*. Artech House, Inc.

Schubert, S., Neubert, P., Pöschmann, J., and Protzel, P. (2019). Circular convolutional neural networks for panoramic images and laser data. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 653–660.

Seliga, T. and Bringi, V. (1976). Potential use of radar differential reflectivity measurements at orthogonal polarizations for measuring precipitation. *Journal of Applied Meteorology*, 15:69–76.

Straka, J. M., Zrnić, D. S., and Ryzhkov, A. V. (2000). Bulk hydrometeor classification and quantification using polarimetric radar data: Synthesis of relations. *Journal of Applied Meteorology*, 39:1341–1372.

Wilson, J. W., Weckwerth, T. M., Vivekanandan, J., Wakimoto, R. M., and Russell, R. W. (1994). Boundary layer clear-air radar echoes: Origin of echoes and accuracy of derived winds. *Journal of Atmospheric and Oceanic Technology*, 11:1184–1206.

Zhang, G. (2016). *Weather Radar Polarimetry*. CRC Press.