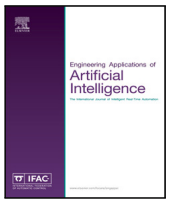
Engineering Applications of Artificial Intelligence 139 (2025) 109543

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

Real-time joint recognition of weather and ground surface conditions by a multi-task deep network

Diego Gragnaniello [a], Antonio Greco [a,*], Carlo Sansone [b], Bruno Vento [b]

[a] Department of Information and Electrical Engineering and Applied Mathematics (DIEM), University of Salerno, Italy
[b] Department of Electrical Engineering and Information Technology (DIETI), University of Napoli Federico II, Italy

ARTICLE INFO

MSC:
0000
1111

Keywords:
Weather recognition
Ground surface classification
Multi-task learning
Deep neural network
Artificial vision

ABSTRACT

Climate change and the occurrence of intense and unexpected weather events highlighted the need for real-time weather warning systems, especially in smart roads and isolated scenarios like rural areas. In this work, we propose to jointly recognize the weather and the ground surface conditions using existing video surveillance systems. Previous works separately tackled these two tasks even if they are correlated to each other. We propose a convolutional neural network with shared weights in the lower layers and two separate classification branches on top to exploit the correlation between the tasks and, at the same time, learn diverse high-level features for each task. Moreover, the network architecture implements attention mechanisms allowing the classification branches to focus on diverse image regions. The method is versatile and allows us to train the network on partially labeled data. The experimental analysis on real data demonstrate the effectiveness of the proposed method on both tasks, confirmed by the accuracy comparison with existing methods for the recognition of weather and ground surface conditions. The multi-task solution improves the inference speed (50 frames per second) and reduces the required memory (less than 1 GB) with respect to a system with two different single-task approaches; these results confirm that the proposed solution is ready for video surveillance applications to support smart cities.

1. Introduction

Nowadays, with the intensification of climatic conditions around the globe, critical circumstances that may pose a threat to people's health are increasingly common. In particular, unpredictable intense precipitations and strong winds are dangerous for both pedestrians and drivers and may damage infrastructures and agriculture, especially in case of flooding (Jun et al., 2024). The need for real-time weather recognition systems, which have a greater impact when employed remotely in isolated rural areas, fostered the research regarding automatic weather recognition systems. Early works proposed automatic systems deployed in static weather stations (Sandnes, 2012) or onboard the vehicles (Qian et al., 2016; Almazan et al., 2016; Onesimu et al., 2021). While the latter solution requires certified hardware to be installed onboard, the former is less expensive due to the reuse of existing infrastructures and provides information about a remote location with no human presence. The advances in Computer Vision made it possible to design video-based weather recognition systems using the surveillance infrastructure (Gragnaniello et al., 2024a). An early work that raised the attention of the scientific community on the potential of automatic weather classification for traffic control using surveillance cameras was presented in Lagorio et al. (2008). Nowadays, this potential is further increased by the presence of low-cost smart cameras. These connected devices, already employed for numerous video-based automatic event detection tasks, can perform real-time detection of the current and local weather conditions to be shared over the network. The widespread of these cameras at low prices for video surveillance set up a vast network infrastructure that is sometimes employed for adverse weather detection, too.

The recent literature mainly distinguishes between two different tasks depending on the goal of the application: recognizing the weather or the ground surface conditions. The former, which regards recognizing the atmospheric conditions (e.g., sunny, foggy, or rainy), is interesting in many application scenarios, and barely any camera can be employed. Meanwhile, the latter is devoted to detecting ground surface conditions (e.g., dry or wet) and is suited for road monitoring applications to detect potential hazards for oncoming drivers. This is particularly true for off-road tracks, like in agricultural fields, where mud or snow may make the track impassable. Considering the two functionalities as a single weather warning system, there are at least three areas in which these features can be profitably used: i) in modern

* Corresponding author.
E-mail address: agreco@unisa.it (A. Greco).

https://doi.org/10.1016/j.engappai.2024.109543
Received 21 January 2024; Received in revised form 4 August 2024; Accepted 22 October 2024

smart roads, to provide real-time feedback on weather and road surface conditions, with the possible aim of adjusting the lighting intensity of streetlights accordingly or to alert the authorities in time; on board of vehicles used for road monitoring, which could geographically localize sections with dangerous weather or asphalt conditions for drivers; in agriculture, to encourage interventions by farmers or autonomous robots for the watering and irrigation of agricultural fields and for the protection of plants from bad weather and frost. For each task, some example images representing the classes considered in our experiments are depicted in Figs. 3 and 4. Obviously, the ground surface is highly influenced by the weather, and the two tasks are correlated with each other. Of course, this correlation can be observed only in a suitable time scale. As an example, the ground is dry when the rain starts, then it gets wet after a period whose duration depends on the terrain characteristics (e.g., the drainage capacity). As a consequence, a real-time system cannot simply infer the ground surface condition from the recognized weather one. Note that less likely conditions, like wet ground on a sunny, are more dangerous, since the driver may experience unexpected low grip.

In this work, we present a novel framework that allows addressing the two classification problems at the same time using a single multi-task neural network. The main contributions of our work are the following. Firstly, we designed a convolutional neural network architecture with shared low-level weights and two separate high-level classification heads. By sharing low network layers, our network exploits the correlation between the two tasks by learning a common low-level representation. At the same time, two different classification branches allow us to learn high-level task-specific features. By doing so, the training procedure produces a single efficient and effective model, exploiting the two task correlations while adapting the classification stages for each of them. Secondly, we can mask missing labels during the training phase. If a training sample lacks the label for a specific task, the corresponding classification branch is not trained while the rest of the network is. This allows us to fully leverage partially labeled datasets or single-task ones, thus greatly enlarging the training set. Thirdly, we carry out both tasks employing one neural network suitable for edge computing, i.e., that is memory and time-efficient and fits on smart surveillance cameras. It should be noted that the proposed approach is highly scalable. Indeed, as Computer Vision techniques improve, it is reasonable to include more classes in the recognition system. For the proposed framework, an increase in the number of classes for one task does not affect the complexity of the other classification branch. At the same time, we can consider the increase in the complexity linear with the number of classes. The main drawback of our framework is the increase in complexity of the training phase. Firstly, if we want to change or add classification tasks, the whole network must be re-trained unless incremental learning is adopted (Kanakis et al., 2020). Moreover, since two losses must be jointly optimized by the proposed architecture, the convergence is harder to reach. The difficulty depends not only on the type of loss functions but also on the training set, which should contain fewer possible mislabels or ambiguous samples. To alleviate the convergence issues, we rely on advanced training procedures that will be described in the paper.

2. Related works

Since the advent of powerful image classification methods, several approaches have been proposed for weather condition recognition. Like for many other computer vision applications, the techniques shifted from handcrafted features, used to train classical machine learning classifiers, to more recent deep learning based approaches. The vast majority of techniques address only one among weather classification and ground surface classification. We carried out the analysis of these approaches in the first paragraph of this section, specifically focusing on the problem definition, the adopted dataset, and the classification technique. More recently, two works addressed both tasks at the same time. Since these are the most related works, we thoroughly analyze the differences between our approach and those already published in the second paragraph. An overview of the discussed methods is summarized in Table 1, which reports for each of them the addressed tasks, the adopted datasets, the number of classes, and the achieved accuracy.

Table 1
Summary of the related works, ordered by year. We first specify if weather (W), ground surface (GS), or both jointly (W+GS) are recognized for each method. Thus, for each of the tasks addressed, we report the dataset used, the number of considered classes (#Classes) and achieved accuracy.

Method	Task	Dataset	#Classes	Acc.(%)
Elhoseiny et al. (2015)	W	Private	2	91.1
Zhang et al. (2016)	W	MWI	4	71.4
Lu et al. (2017)	W	TCWD	2	91.4
Kang et al. (2018)	W	MWI	4	92.0
Nolte et al. (2018)	GS	Private	6	84.7
Zhao et al. (2019)	W	TCWD FCWD	2 5	94.9 83.4
Al-Haija et al. (2020)	W	MCWRD	4	98.2
Carrillo et al. (2020)	GS	RWIS	3	91.0
Grabowski (2020)	GS	RoadCCTV	3	96.0
Khan and Ahmed (2022)	W GS	Private Private	3 3	97.3 99.1
Xie et al. (2022)	W	FCWD	5	85.6
Abdelraouf et al. (2022)	W+GS	Private	4	91.3 93.7
Mittal and Sangwan (2023)	W	MCWRD	4	97.8
Chen et al. (2023)	W	Private	5	88.9
Zhao et al. (2023)	GS	RCSD	27	97.5
Samo et al. (2023)	W+GS	RWD	7	91.9 91.3
Gragnaniello et al. (2024b)	W GS	FCWD Private	5 3	87.8 83.8

2.1. Single task approaches

Several methods have been proposed to either classify the weather or the ground surface from CCTV cameras. These two classification tasks exhibit different challenges, which moreover may significantly vary depending on the considered scenario or weather conditions. Due to this, in the following, we separately discuss the techniques proposed for weather classification and ground surface recognition.

2.1.1. Weather

Weather classification is, among the two, the task that first gained more attention from a large part of the scientific literature. Early works focused on a binary problem, e.g., sunny versus rainy classification, mainly using handcrafted features It was in the last decade that several data-driven approaches were proposed. Their superior discriminative power allows them to outperform classical approaches on both binary and multi-class problems. As an example, in Elhoseiny et al. (2015), the authors address the two class problem (i.e., sunny versus cloudy) using the AlexNet architecture. By analyzing the impact of each layer in this rather shallow architecture, they show that low-level features capturing spatial information are important for weather classification. The same problem is addressed by Lu et al. (2017), who collected a two-class dataset TCWD, made available to the community. In their paper, the authors propose to concatenate features extracted by AlexNet, after finetuning for the binary classification problem, with handcrafted binary features representing the presence or absence of weather clues like shadows or haze. An SVM model trained with these features reached an accuracy of 91.4%. Later, in Zhao et al. (2019), the

same classification problem is addressed by automatically extracting weather clues through a CNN. A multi-task architecture is employed to simultaneously classify the weather category (sunny versus cloudy problem) while performing weather-cues semantic segmentation, like for example the blue/dark sky or clouds. To this aim, the authors extended the dataset previously presented in Lu et al. (2017) by manually depicting a bounding box around each weather clue. The auxiliary segmentation task helps the network to better distinguish between the two weather classes, improving the classification performance up to 94.9%.

In the same work, the authors presented a novel five class weather dataset (FCWD) comprising samples of Sunny, Cloudy, Rainy, Snowy, and Fog. Multi-class weather category recognition problems are challenging due to the imbalance among class samples. Moreover, different weather clues can be contemporary present in the same sample, misleading the classifier. In Zhao et al. (2019), the authors reported an accuracy of 83.4% on the five-class dataset, with the Cloudy and Fog classes being the most challenging ones. A similar approach has been presented in Xie et al. (2022). The authors improved the segmentation branch, proposing a multi-resolution architecture, and adopting a mechanism to dynamically adapt the weights of the classification and segmentation losses. The accuracy achieved on the same two-class and five-class datasets presented in Zhao et al. (2019) are 96.3% and 85.6%, respectively. The five-class problem is addressed in the recent work (Gragnaniello et al., 2024b), where the authors show that the efficient MobileNet-V2 architecture can be trained to separately address either weather and ground surface, but optimizing hyperparameters like the input resolution. The technique can run in real-time on embedded devices, but requires cropping the region of interest for each task. Among the first works addressing a multi-class weather classification problem from a single image, in Zhang et al. (2016) the authors collected a large dataset, namely Multi-class Weather Image (MWI), comprising 20k images belonging to the 4 classes Sunny, Rainy, Snowy, and Haze. The method was based on both local features, designed to extract specific weather clues from sky or shadow and computed using the Histograms of Oriented Gradients (Dalal and Triggs, 2005), and global features like contrast and saturation to exploit the overall illumination condition. Using a classification model based on dictionary learning, the author surpassed other approaches like SVM, reaching an overall accuracy of 71.4% on the MWI test set. The very same problem is tackled in the work of Kang et al. (2018). The authors proposed to fine-tune a CNN borrowed from object recognition applications and pre-trained using the ImageNet dataset (Deng et al., 2009). The network was trained to recognize one of the three adverse weather conditions (i.e., Rainy, Snowy, and Haze). In the absence of detections, the Sunny class was predicted. The experiments carried out with the AlexNet and GoogleNet architectures, demonstrated their superior performance, with the latter reaching an accuracy of 92.0%. Similar approaches have been pursued by Al-Haija et al. (2020), Khan and Ahmed (2022) using the more lightweight ResNet18 architecture. All these experiments are conducted on proprietary datasets to recognize a different set of weather categories, like Sunrise and Shine, which prevents a direct performance comparison. Similar classes are considered in Gbeminiyi (2018), which includes 1125 images divided into four categories, i.e., cloudy, rainy, shiny, and sunrise (MCWRD). Mittal and Sangwan (2023) adopted this dataset to conduct an experimental assessment of Computer Vision CNNs fine-tuned for the weather. Among them, InceptionV3 trained with Logistic Regression ranked first, however the significance of this result is limited by the small number of dataset samples. The focus of this work is on the scalability of the training procedure, which is implemented using the Spark platform to be efficient when using huge datasets. In Chen et al. (2023) the authors propose MASK-Convolutional Neural Network-Transformer (MASK-CT), a method for multi-class weather recognition that combines CNNs and Vision Transformers to extract effective features for weather classification. They enhance the generalization capability of their approach by adopting an augmentation strategy that randomly masks part of the image or of the labels during training. MASK-CT has been tested on three sets of images acquired in real-time, achieving an average accuracy (overall recall) equal to 88.9% in the recognition of five weather categories.

Generic weather category recognition is very useful for low-cost constant monitoring of weather using video surveillance cameras. As a drawback, the performance of the classifiers varies a lot on the scene acquired (e.g., the rain appears very different when framed in a near scene or a landscape). From the point of view of the application scenario here considered, these methods provide useful information about driving visibility, however, ground surface conditions must be inferred by the driver. Indeed, the same weather can yield different ground surface statuses. As an example, rainy weather can be associated with either a mostly dry or flooded ground depending on the rain intensity and duration, and the soil material. Thus, two systems are needed to provide the weather and the ground surface conditions. Since these tasks correlate with each other, two separate systems would be inefficient. A multi-task approach would exploit this correlation to reliably and efficiently provide the weather and ground surface conditions at the same time.

2.1.2. Ground surface

Ground surface condition recognition from video surveillance cameras is a task that gained more attention only in recent years. Early works have tested the use of CNNs to address this task. Nolte et al. (2018) revised the available datasets containing road images, most of which acquired using onboard cameras, and selected a subset of the classes to perform the network training and validation. Among the considered architectures, ResNet50 performed best. The performance was further increased by adding training samples, crawled from the web, belonging to the minority classes. More recently, a significant effort was made to collect, label, and make available datasets specifically for the road condition detection task from fixed cameras. The work by Carrillo et al. (2020) presents an analysis focused on snow level recognition on the road surface. The dataset includes images and weather data (e.g., humidity, pressure, and so on) acquired by several stations in Canada and labeled considering road status only. The authors conducted two parallel experimental analyses. On one hand, they trained seven different CNN architectures to classify the snow level from acquired images. On the other hand, the authors tacked the same task by training three machine learning classifiers, namely SVM, RF, and NB, using the auxiliary weather data. The reported performance exhibits the potential of both methods and the overfitting threat when using deep neural networks with many trainable parameters. A similar work has been presented by Grabowski (2020). In this case, the authors collected a dataset composed of images acquired by weather stations in Poland and labeled with both the precipitation type/level and ground surface category. Six precipitation classes are considered, namely no precipitation (the vast majority class), dew, continuous, intense, shower, and snow. Instead, the seven road conditions collected using the stations' weather sensors are dry (the majority class), moist, saline, wet, snow, rime ice, and ice. The experimental, however, was limited to the dry, wet, and snow classes due to the insufficient number of samples of the remaining ones and to allow for performance comparisons with those reported in previous works. All the selected CNN architectures reached 96% accuracy, with DenseNet (Huang et al., 2017) surpassing the others by little. Residual networks architecture was employed by Khan and Ahmed (2022), too. In this work, the same approach has been adopted to recognize the weather category and the ground surface condition, separately. In both cases, fine-tuning the ResNet18 architecture pre-trained on the ImageNet dataset (Deng et al., 2009) performed better than the other techniques selected for performance comparison on the proprietary dataset. In Gragnaniello et al. (2024b) the same problem is addressed utilizing a more recent architecture, namely MobileNet-V2, implementing separable convolutions and capable of being executed on

devices with limited resources. The same architecture is used for either the ground surface and weather, by separately optimizing the training hyperparameters. A very recent work (Zhao et al., 2023) collected the Road Surface Classification Dataset (RSCD), composed of image patches of road surface video frames. Each patch is labeled according to three different ground characteristics, namely the road material (e.g., asphalt, concrete, mud, etc.), the unevenness level (smooth, slight or severe), and the friction level that represents the road weather category. The considered categories are Dry, Wet, Water, Fresh Snow, and Melted Snow. The proposed method, based on the EfficientNet (Tan and Le, 2019) architecture, tackles the multi-task classification as a single-task classification in the set of classes obtained by all the class triplets combinations. While simplifying the CNN classification stage and the training procedure, the main drawbacks of this approach are the extraction of a single feature representation for all the tasks and the scalability with respect to the number of classes. Section 2.2 will better discuss these aspects in relation to similar approaches proposed to address the multi-task weather and ground surface classification.

Data-driven approaches for ground surface condition recognition are powerful but difficult to control. Indeed, when the dataset is not carefully labeled according to the real ground conditions solely, the network can be biased by weather clues that may not correspond to the ground condition. For example, the network could predict the dry ground surface class in the presence of sun or, vice-versa, predict the wet ground surface class when it is cloudy. To prevent this, samples for which the weather and the ground surface conditions differ, e.g. wet ground on a sunny day, are very useful. Unfortunately, they are the minority of the dataset. Multi-task approaches can alleviate this risk by explicitly addressing both weather and ground surface condition recognition at the same time.

2.2. Multi-task approaches

To the best of our knowledge, this is the first attempt to address this joint classification problem using a multi-task network. Very recently, two works (Abdelraouf et al., 2022; Samo et al., 2023) proposed methods to simultaneously recognize the ground surface and weather conditions by means of a Visual image Transformer (ViT) (Dosovitskiy et al., 2020). In Samo et al. (2023), onboard camera images are classified either in a ground surface or a weather category from Clear, Sunny, Cloudy, Wet, Snowy, Rainy, and Foggy. The dataset, namely Road Weather Dataset (RWD), has been collected and shared by the authors. Each image in the dataset, extracted by online video, was manually labeled by visual inspection, i.e. without the help of additional information acquired by specific sensors. To get rid of the dataset unbalance among the seven classes, the focal loss is employed during the training phase: this loss function allows to dynamically scale the weights of the correctly classified samples belonging to the most represented categories and, thus, to focus the learning on the misclassified samples of the less represented classes. Reformulating the multi-task problem as a single task significantly simplifies the model complexity in terms of both architecture, which needs a single classification stage, and training procedure that optimizes a single loss function. However, this approach has the disadvantage that weather and ground surface conditions cannot be both recognized when they are present in the image at the same time. Training the network to select one of them can mislead the procedure and lead to bias. This problem is solved in Abdelraouf et al. (2022), where the authors train the neural network to recognize any combination of classes of the two tasks. Thus, the approach tackles the two tasks as a single one whose classes are all pairs of classes of the two single tasks. Eventually, the network outputs are combined to obtain predicted probabilities for each task. As for the previous method, the training procedure is simple and robust. On the other hand, this approach still has some drawbacks. First, just like the previously discussed method, it does not decouple the learning process for each task, thus preventing the specialization of high-level features for weather and ground surface classification, separately. This is even more important considering that clues to tackle these two classification tasks are spatially separated in the image. However, it is worth noting that the authors of both works (Abdelraouf et al., 2022; Samo et al., 2023) alleviated this by leveraging on the spatial self-attention mechanism implemented by ViT. Another important consequence of these approaches is that each sample must be labeled for both tasks, which limits data suitable for the training phase. Due to the scarcity of available labeled data, it could be useful to have a technique able to handle missing labels. Second, the proposed combined single-task strategy has limited scalability with the increase in the number of classes. When a finer classification is needed for one task, the problem to address becomes much more complex since the amount of class combinations increases. To alleviate this, approximations can be tacked. As an example, in Abdelraouf et al. (2022) the authors impose that Rainy weather always implies Wet roads, which reduces the problem's classes. Even if this is reasonable, it neglects minority cases and may bias the training when used extensively on several class couples.

3. Proposed method

In this section, we present details about the proposed approach, whose architecture is depicted in Fig. 1. The first and second paragraphs introduce the proposed neural network architecture, presenting the adopted backbone and the introduced modifications to address the multi-task problem. Thus, the third one describes how the neural network is trained. The last paragraph presents the collected dataset, composed of online available data, partially labeled for the purpose of this work.

3.1. Backbone for feature extraction

The backbone of the proposed multi-task architecture is inspired to ConvNeXt (Liu et al., 2022). This CNN has been developed starting from the ResNet architecture to implement the most effective features of Transformer architectures (Liu et al., 2021). The result is a network exploiting all the CNN peculiarities in terms of priors and efficiency while surpassing most of the recently proposed Transformer architectures on classification tasks. In doing this, the authors replace the costly self-attention mechanism, peculiar to Transformer architectures, with very efficient tricks such as larger convolutional kernels followed by point-wise convolutional stages. At the same time, the authors state that a cross-attention module would be desirable when dealing with multi-modal input. Inspired by the impressive results achieved by this neural network, we adapted the ConvNeXt architecture to address the considered multi-task problem as described in the next paragraph.

To let the network focus on specific image areas, which can be different for the two tasks, we integrated an attention module in the original ConvNeXt architecture, namely the Convolutional Block Attention Module (CBAM) presented by Woo et al. (2018). This technique has been selected among the variety of techniques proposed in the literature since it implements two attention mechanisms, i.e., spatial-wise and channel-wise, which fit our needs. Two CBAM modules are trained, together with the whole network, to provide a set of spatial and channel weights given the low-level feature map extracted by the common network branch. These sets of weights, different for each classification head, are used to focus the network attention on the image region and feature channel peculiar to each task.

3.2. Multi-task network architecture

Multi-task neural networks can effectively exploit the tasks' correlation while efficiently predicting sample classes using a single network architecture (Foggia et al., 2023). To this aim, we modified the ConvNeXt architecture to produce two outputs satisfying the two-task

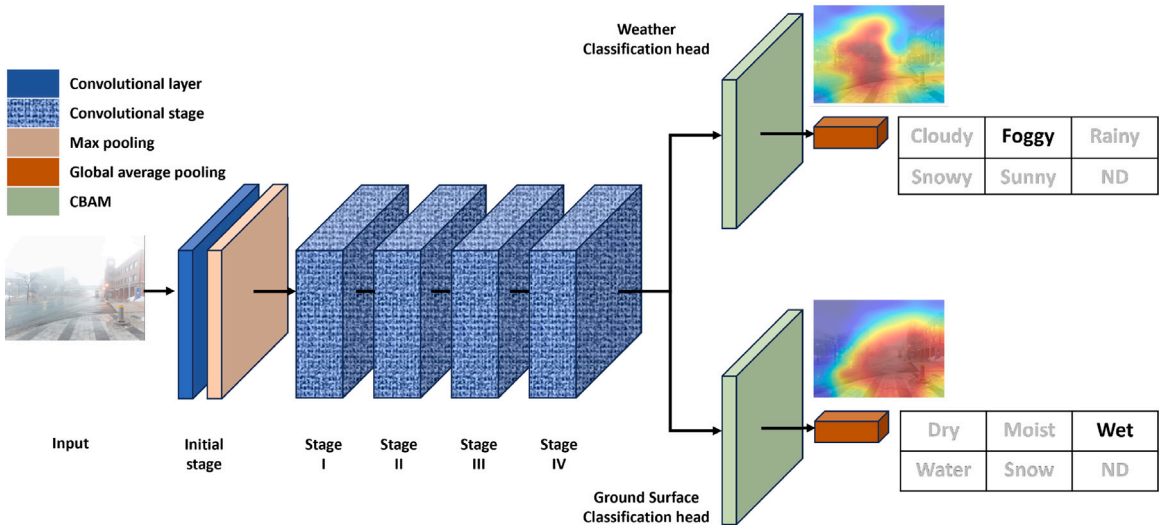


Fig. 1. Architecture of the proposed multi-task neural network. The backbone for extracting shared low level features is based on ConvNeXt, while two task-specific attention modules based on CBAM constitute the weather and ground surface classification branches. The neural network is trained with a masked asymmetric loss (M-ASL), that allows to effectively learn representative features even in case of missing labels and unbalanced datasets, while GradNorm fosters a balanced learning across the two tasks.

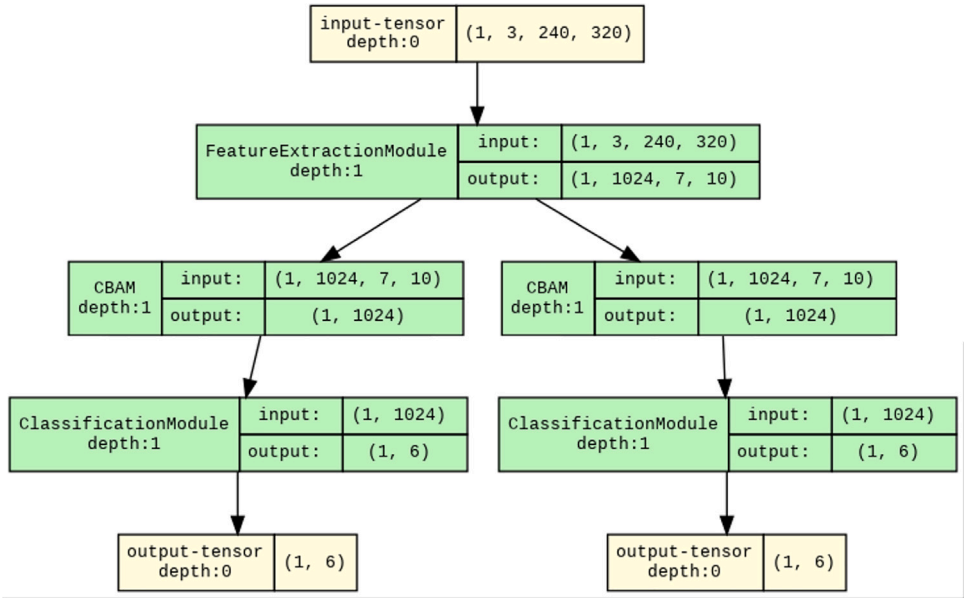


Fig. 2. Details of the proposed neural network architecture based on ConvNext-base model and modified to include two classification heads, each equipped with an attention module.

problem presented in this work. As can be observed in Fig. 1, the lower part (i.e., the first layers) of the backbone remains unchanged to extract a common low-level representation of the input image, meanwhile we modified the top layers. The same feature map is provided as input to two different network heads, each replicating the top backbone layers. These are the network branches performing the weather and the ground surface classification tasks. This helps to decouple the two tasks, making our approach scalable when the number of classes increases while exploiting the intrinsic correlation between them by building the decision of both tasks on top of a common set of low-level features.

The feature map that feeds the two branches has been selected to satisfy two needs. On one hand, it must be at a sufficiently high level to represent meaningful semantic information, like the presence of clouds or water on the ground. On the other hand, it should keep the spatial information of the extracted features, which is lost at higher levels of CNNs after global pooling operations. In particular, after the bottom layers of the network which remain unchanged, allowing us to exploit the pretrained weights of ConvNext, the feature map spatial dimension is equal to 7×10 whereas the number of feature maps is 1024. This feature map is fed to the two classification branches. Each spatial attention module performs a different weighted average of the adjacent pixels in the feature map, allowing the two classification branches to focus on different image regions. Each weighted average yields a 1024-dimensional feature vector, which is finally processed by the corresponding classification head. This consists of 4 consecutive fully connected layers interleaved by ReLU activations and dropout layers. Fig. 2 highlights the details of the proposed architecture in terms of input and output dimensions for an input image of resolution 320×240 pixels; we demonstrate in Section 4.2 that this resolution is enough to achieve the best results.

3.3. Learning procedure

An important consequence of the proposed network architecture is its ability to exploit partially labeled data during training. Indeed, samples associated with just one of the two labels are processed by the classification head for which the annotation is available, producing one output instead of two. Thanks to a masked loss function, during the backward pass, only the network layers involved in the forward pass and loss computation are traversed. This allowed us to greatly increase the amount of training data, thus reducing the risk of overfitting.

To train the network, we adopt the Asymmetric Loss (ASL) framework proposed by Ridnik et al. (2021) with the aim of balancing the number of positive and negative labels in multi-label classification problems. This loss function assigns different weights to positive and negative samples for each class of interest; by appropriately configuring these weights, we are able to balance the learning with respect to the a priori distribution of the dataset and adapt the response of the neural network to the expected distribution in the real application. Formally, the ASL of one sample and a generic task is articulated as:

$$\mathcal{L}_{ASL} = \sum_{i=1}^{n} \left[-y_i L_+ - \left(1 - y_i\right) L_- \right], \quad \text{with} \begin{cases} L_+ = (1 - p_i)^{\gamma_i^+} log(p_i) \\ L_- = p_i^{\gamma_i^-} log(1 - p_i) \end{cases} \tag{1}$$

where n is the number of task classes, while y_i and p_i respectively represent the label and prediction probability of the ith class in the considered image. Similarly, γ_i^+ and γ_i^-, namely the positive and negative focusing parameters, respectively, regulate the importance assigned to either the ith class for positive or the negative samples. With respect to the original ASL formulation, we do not perform any probability shift, thus avoiding discarding any negative samples even when they are easy to classify. This is motivated by the limited number of classes of our problem with respect to the hundreds of classes of object detection benchmarks, for which ASL was originally proposed: for both tasks, the number of attributes is equal to 6. Since we address a two-task problem, the resulting loss function is the sum of two ASL terms (1):

$$\mathcal{L} = \mathcal{L}_{ASL}^{W} + \mathcal{L}_{ASL}^{G} \tag{2}$$

The first term, namely $\mathcal{L}_{ASL}^{W}$ accounts for the weather classification task, while the second one, i.e. $\mathcal{L}_{ASL}^{G}$ is used to learn the ground surface classification. To set the focusing parameters of both terms, we followed the guidelines suggested in Ridnik et al. (2021) configuring the γ_i^+ weights to 0 and varying the γ_i^- values according to the distribution of the classes in the training set and reported in Table 2. Thus, for the weather classification task, we choose γ^- equal to 4 for snow, which is the less represented class, 1 for cloud and 2 for the others. Similarly, for the ground surface recognition task we set γ^- equal to 4 for snow, 2 for wet, water, and moist, and 1 for the others.

To balance the training in terms of task and class representativeness, we employ two strategies. As regards the tasks, we resort to GradNorm (Chen et al., 2018), a technique that dynamically balances the magnitude of the gradients back propagated from each of the loss functions of a multi-task network during training. This is achieved by introducing one additional trainable parameter for each term in the loss function (2), which results:

$$\mathcal{L} = w^W \mathcal{L}_{ASL}^{W} + w^G \mathcal{L}_{ASL}^{G} \tag{3}$$

where w^W and w^G, namely the weights of the two tasks' loss terms, are optimized during training. In order to exploit all the samples in our training dataset, we include in the minibatch even those labeled only for one task. For those samples, the loss, and thus the backpropagation, is computed only for the task for which the label is available, while the other is masked. We refer to this as masked asymmetric loss (M-ASL). As for the class balance, since real datasets are unbalanced, we wisely compose the mini-batch to approximately achieve the same amount of samples for each class of both tasks.

For what concerns the optimizer, we use AdamW with a weight decay equal to 0.01 and an initial learning rate set to 10^{-5}. The learning scheduler is based on cosine annealing, with a final learning rate set to 10^{-7}. We implement an early stopping procedure with patience equal to 5: if the average accuracy among the two tasks on the validation set does not improve for 5 consecutive epochs, the training procedure stops. By using this learning procedure, the training ended for all the experiments in less than 20 epochs, ensuring a fast convergence. No data augmentation has been done, since it did not improve the performance, as demonstrated in Section 4.2.

3.4. Dataset

No standard classes have been defined in the literature for weather classification and for recognizing ground surface conditions. Therefore, it is primarily necessary to define the classes considered for the two tasks of interest. As regards the weather, we chose the most used categories in the literature, which cover more or less all possible meteorological situations: cloud, fog, rain, snow, sun. In addition, to handle situations of uncertainty or those images in which the sky is not visible, we considered an additional undefined class (ND, not defined), which allows the trained systems to reject uncertain samples for the weather recognition task. Examples of images and their corresponding weather categories are reported in Fig. 3.

As for the ground surface conditions, in addition to the standard dry, moist, snow, and wet categories, we consider also a water class, which includes all those images depicting flooded land or bodies of water. Again, we considered a rejection class (ND) for those samples that are uncertain or where the ground is not visible. Samples annotated with their corresponding ground surface categories are depicted in Fig. 4.

Since there are no datasets annotated with both labels and with the considered categories, for our experiments we decided to rely both on publicly available datasets and on new images collected and manually labeled for the two tasks of interest. In particular, we gathered from YouTube more than 3000 video clips of around 200 live webcams all over the world. These scenarios were selected among the numerous available online due to their complex and heterogeneous operating conditions. Both rural and urban scenarios were considered, with variable camera position, illumination conditions and density of passages by people and vehicles, in order to have samples with different features and possible partial occlusions. The recordings were made by selecting the moments depicting different weather and ground surface conditions. Each video clip lasts approximately 10–15 s and the sequences of images were acquired from each webcam at different times of the day. In this way, we are able to increase the diversity, the size and the representativeness of the dataset. One image per second was extracted from the videos and all these samples were annotated with the weather and the ground surface categories. Moreover, we have selected 12,679 samples from the Five Class Weather Dataset (Gao, 2019), annotated with weather labels, and 26,935 from the Road CCTV Dataset (Grabowski, 2020), annotated with ground surface labels, to complete the training set of our method. The selection of these samples was carried out by choosing those that frame both the ground and the sky, as we are interested in this kind of images for the two tasks; furthermore, we chose more samples from the less represented classes, in order to balance, at least partially, the training set. These samples are annotated with only one of the labels, but this is not a problem since the learning procedure we propose is designed to handle this aspect. The details about the 110,634 samples of the training set are reported in Table 2.

Then, we composed the validation and the test sets by annotating the missing labels of around 1500 samples (for both sets) selected from the Five Class Weather Dataset and the Road CCTV Dataset, not used in the training set. In the validation set, we have considered the ND

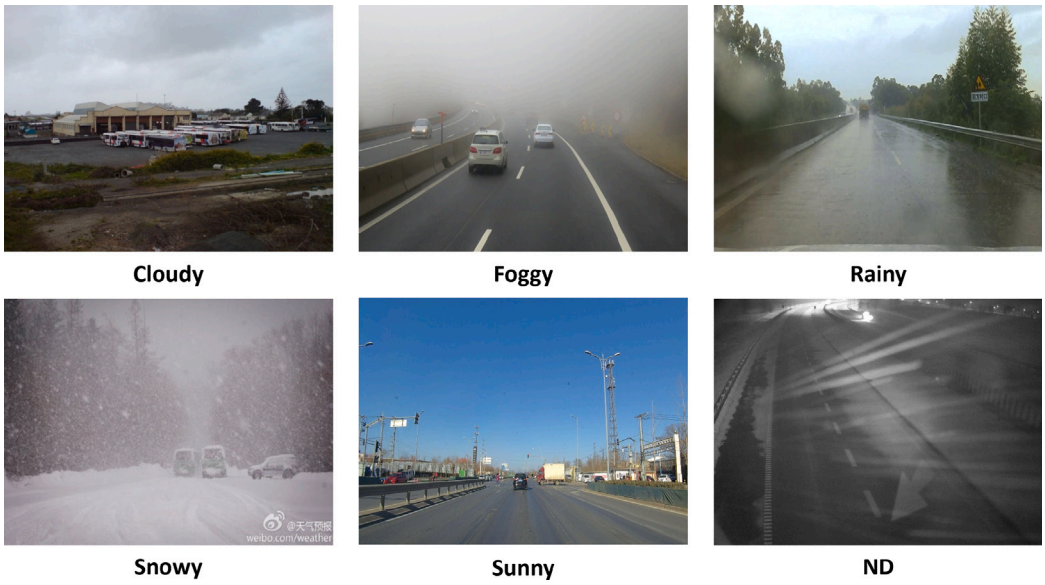


Fig. 3. Image samples for the considered weather categories.

Fig. 4. Image samples for the considered ground surface categories.

Table 2
Class distribution of the training samples for each task addressed in our experiments. The last two columns report the number of samples to reject (ND) and those with missing labels (ML), respectively.

Weather	Cloudy	Foggy	Rainy	Snowy	Sunny	ND	ML
	35,997	5,430	5,554	2,152	18,842	15,724	26,935
Ground surface	Wet	Dry	Water	Snow	Moist	ND	ML
	17,735	22,347	15,751	2,066	12,687	27,369	12,679

class for both tasks, while we have excluded the reject category from the test set. The latter frames a different scenario in each image and was built ensuring a fairly balanced distribution between the samples of the various classes for both tasks. Also this set contains samples that are quite variable in terms of framed environment, camera positioning, occlusions of people and vehicles, and lighting conditions. More details about the composition of the test set are reported in Table 3.

4. Results

In this section, we present the experimental setting and discuss the obtained results. In our analysis, we first motivate the use of a single multi-task network by implementing the comparison with two single-task baselines implementing the very same backbone neural network and trained with similar hyperparameters. On one hand, this allows

Table 3
Class distribution of the test samples for each task addressed in our experiments. In this set, there are no samples to reject (ND) or those with missing labels (ML).

Weather	Cloudy	Foggy	Rainy	Snowy	Sunny
	309	257	325	225	298
Ground surface	Wet	Dry	Water	Snow	Moist
	221	205	154	185	174

Table 4
Results achieved by the proposed multi-task neural network (MT) in terms of accuracy (%), FPS and required memory (MB), compared with the corresponding single-task baselines (ST-W and ST-G). The experiments have been done on an NVIDIA Quadro RTX 8000 GPU.

Method	W(%)	G(%)	Avg.(%)	FPS	Memory (MB)
ST-W	87.27	–	67.65	36	1,234
ST-G	–	48.03			
MT	**89.32**	**60.70**	**75.01**	**50**	**818**

us to weigh the effectiveness of our approach, which exploits the tasks' correlation to improve performance. On the other hand, from an application point of view, this experiment compares two systems, i.e., the single- and multi-task, requiring quite different resources in terms of time and memory space of the host device. After that, we present our ablation study, whose results support from an experimental point of view the choices made in the method design. In particular, we present the gap between the performance achieved by our best solution and several variants obtained by changing one hyperparameter at a time. Finally, we compare the performance of our proposal with that of existing methods for weather and/or ground surface recognition. This allows us to better assess the performance of our method on the adopted dataset.

4.1. Multi-task vs single-task

The baselines of our work are two ConvNeXt architectures trained to address weather classification (W) and ground surface recognition (G), separately. These networks are trained using only the subset of the training set labeled for the corresponding task. All the hyperparameters are the same as the proposed multi-task approach. The results achieved by the proposed multi-task solution (MT), compared with the single task baselines (ST-W and ST-G), are reported in Table 4.

The proposed solution outperforms the single-task counterparts over all the performance metrics, demonstrating better effectiveness and efficiency. Our multi-task neural network achieves 89.26% weather classification accuracy and 60.70% ground surface recognition accuracy, which is better than that obtained by the single task baselines that is 87.27% for weather classification and 48.03% for ground surface recognition. While the results are quite similar in terms of weather classification, we can note a significant improvement in the recognition of ground surface conditions. In our opinion, the proposed approach exploits the tasks' correlation, improving the performance on both tasks but, in particular, on the most challenging one, i.e., the ground surface recognition capability. As for the needed resources, the proposed multi-task neural network is faster than the two single-task counterparts when these are executed sequentially (50 vs 36 FPS) and requires less memory space when they are executed in parallel (818 vs 1234 MB). Therefore, the experimental results confirm the validity of the proposed solution in terms of accuracy and computational resources.

To better characterize the performance achieved by the proposed multi-task neural network on the various classes, we computed the test set normalized confusion matrices, reported in Fig. 5.

As for the weather classification (see Fig. 5, left), we can observe that the correct classification rate is similar among classes, ranging from 84.1% of the cloudy class to the 94.2% of the foggy one. Misclassifications mostly occur between cloudy, sunny, and rainy classes.

Table 5
Results achieved by the proposed multi-task neural network (top), compared with variants obtained by changing the attention mechanism (Att.), the loss function, the learning rate (lr), the input resolution (Res.), and by using data augmentation (Aug).

Att.	Loss	lr	Res.	Aug.	W(%)	G(%)	Avg.(%)
CBAM	M-ASL	1e−5	320 × 240	None	89.32	60.70	**75.01**
HAM	M-ASL	1e−5	320 × 240	None	88.05	59.32	73.68
CBAM	*ASL*	1e−5	320 × 240	None	85.50	**61.53**	73.51
CBAM	*CE*	1e−5	320 × 240	None	88.26	59.11	73.68
CBAM	M-ASL	*1e-4*	320 × 240	None	86.49	51.56	69.02
CBAM	M-ASL	1e−5	640 × 480	None	**89.46**	53.46	71.46
CBAM	M-ASL	1e−5	320 × 240	*Yes*	85.22	54.03	69.62

The weather classification head rarely enables the reject option (i.e. ND class prediction) and only for the samples of the most challenging classes, that are cloudy and rainy. A different behavior is observed for the ground surface recognition (Fig. 5, right). The confusion matrix highlights that the low overall accuracy is not uniformly distributed between classes, that can be grouped in three clusters. Firstly, moist is a very challenging class, whose samples correctly classified only 27.6% times. We manually reviewed them to confirm that they were barely recognizable. The kind of errors confirms this, indeed moist samples are erroneously classified as wet or dry, which are the more closely related classes to the moist one. Secondly, dry, wet, and water classes are challenging. The reported correct classification rate is above 50%. The multi-task network probably achieves higher performance on the wet class (72.4%) by exploiting the two tasks' correlation. While dry and wet classes are confused with each other and with the moist one, the water class is often confused with the wet since the network is unable to distinguish the water level from a static image. Lastly, the third case is the snow class for which a 90.8% correct recognition rate is achieved, which seems easy to recognize, possibly due to its color. The ground surface recognition head rejects more samples than the weather counterpart; this is a further confirmation that this task is substantially more challenging.

4.2. Ablation study

To demonstrate the validity of the design choices done for our multi-task solution, we performed various experiments. Firstly, we evaluated the impact of the attention module and the loss function. As for the attention mechanism, we compared CBAM with a similar approach recently proposed, namely Hybrid Attention Module (HAM) (Li et al., 2022). In terms of loss function, we experimented by replacing the proposed masked asymmetric loss (M-ASL) with the vanilla version (ASL), in which only the samples labeled for both tasks are used, and the Cross-Entropy loss (CE). Also, we have trained the network at varying of hyperparameters like the learning rate or with a standard data augmentation pipeline to select one among horizontal flipping, perspective distortion, rotation, and color jitter. Finally, we performed a test doubling the input size from 320 × 240 to 640 × 480. In Table 5 we report the results of the most significant experiments obtained by varying just one hyperparameter at once with respect to the best solution, reported on top of the table.

We can observe that the choice of CBAM instead of HAM as attention mechanism allows to obtain an improvement for both the tasks (89.26% vs 88.05% in weather classification and 60.70% vs 59.32% in ground surface recognition). The same can be noted for the loss function, since the adoption of the M-ASL allows us to improve the performance in both tasks (89.26% vs 88.26% in weather classification and 60.70% vs 59.11% in ground surface recognition) with respect to the CE. On the average, a similar gap is observed when comparing the results with the vanilla ASL, demonstrating the effectiveness of the proposed framework. As regards the learning rate, we used a quite low starting value to preserve the pre-trained weights. Indeed, using a learning rate only ten times higher dramatically reduces the performance

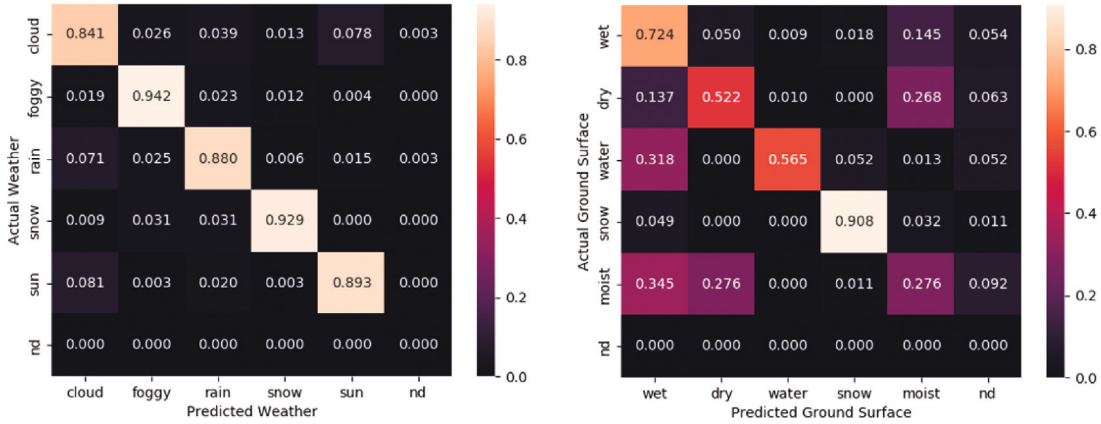


Fig. 5. Confusion matrices of the proposed multi-task neural network on the considered classes for weather classification and ground surface recognition.

Table 6
Comparison of weather recognition results with existing methods. The results marked with * are tested on different samples of the Five Class Weather Dataset, as reported in Xie et al. (2022).

Method	Cloud	Foggy	Rain	Snow	Sun	Avg.
Our	**84.14**	**94.16**	88.00	92.89	89.26	**89.32**
Gragnaniello et al. (2024b)	80.10	93.40	84.30	93.20	**90.30**	87.80
Xie et al. (2022)*	76.70	83.90	**88.60**	**94.40**	87.70	85.60
Zhao et al. (2019)*	73.40	85.00	83.90	92.10	85.40	83.40
Elhoseiny et al. (2015)*	75.30	72.70	79.10	87.60	79.80	79.10
Lu et al. (2017)*	64.80	67.60	76.90	70.20	88.50	76.20

on both tasks, yielding the worst result. Even doubling the input size does not provide advantages in terms of overall accuracy; this result is probably due to the fact that the dataset is not large enough to perform an effective learning with a neural network with such many weights. Therefore, the chosen resolution (320 × 240) makes the most of the potential of the training set and allows reducing both processing time and memory occupation. Similarly, a standard augmentation is unable to produce useful samples. Perhaps ad-hoc geometric transformations should be designed for applications from static cameras. At the same time, color transformations could affect weather recognition, distorting useful information.

4.3. Comparison with existing methods

Comparing our method with others has been challenging because most authors do not publish the code used for training and inference of their proposed neural networks, or the weights learned during training. Therefore, we had to find a compromise for both tasks to compare our results with those of other methods proposed in the literature. The comparison with other weather recognition approaches is reported in Table 6.

Except for the method proposed in Gragnaniello et al. (2024b), that has been trained on the same 5 classes and evaluated on the same test set, the other approaches have been tested on different images of the Five Class Weather Dataset; since the software of these approaches is not available, we refer to the work of Xie et al. (2022) for the performance comparison. We can observe that the proposed multi-task neural network outperforms all the counterparts, that are single-task neural networks specifically trained for the five weather classes of interest. The 89.32% average accuracy is 1.5 percentage points higher than the one achieved by the approach presented in Gragnaniello et al. (2024b), around 4 percentage points better than the algorithm described in Xie et al. (2022) and 6, 10 and 13 percentage points greater than the other methods (Zhao et al., 2019; Elhoseiny et al., 2015; Lu et al., 2017). This result demonstrates the effectiveness of the proposed solution in weather recognition.

Table 7
Comparison of ground surface recognition results with Gragnaniello et al. (2024b) on their test set.

Method	Dry	No Dry	Flooding	Avg.
Our	82.10	94.06	71.64	87.35
Gragnaniello et al. (2024b)	82.80	84.70	99.80	83.80

Since there are no publicly available methods for the recognition of ground surface conditions that have been trained to predict our classes of interest, we applied our multi-task neural network on the dataset used in Gragnaniello et al. (2024b), which considers three classes: dry, no dry and flooding. The only class in common with our set is dry, while flooding can be assimilated to our water; therefore, we associated the predictions for the wet, snow and moist categories to the no dry class. With this mapping, we compared the results with the ones obtained by the method proposed in Gragnaniello et al. (2024b), shown in Table 7.

We can observe that, although our method is trained on more classes, it is able to obtain an average accuracy higher by about 4 percentage points (87.35 vs 83.80). The clearest superiority is on the no dry class (94.06 vs 84.70), while on the dry category the performances are comparable (82.10 vs 82.80). On the flooding class the method proposed in Gragnaniello et al. (2024b) is specialized, while in our case the water category includes more situations including bodies of water; for this reason we notice a performance drop on this class (71.64 vs 99.80). In any case, the multi-task neural network demonstrated good generalization capability across a different dataset.

Finally, to provide hints about the interpretation of the network decision process, we computed the class activation maps (CAMs) for each classification branch, separately. A few examples are depicted in Fig. 6. By analyzing the heatmaps, one can pinpoint the image regions on which each classification head focused during inference. As an example, the weather head (Fig. 6, center) focuses on the center of the image but extends toward the top of it in the presence of fog (top-center). Meanwhile, the ground surface head (Fig. 6, right) mostly focuses on the bottom part of the image, thus different behaviors should warn the user of possible unreliable predictions.

5. Discussion

The experimental results demonstrate the effectiveness of the multi-task solution for weather and ground surface classification. The joint learning of both tasks of interest allows to exploit the interdependencies among the classification problems and to obtain better performance with the multi-task network compared to the single-task baselines. The use of GradNorm fosters a balanced learning over the two tasks, ensuring a uniform accuracy improvement for weather and ground

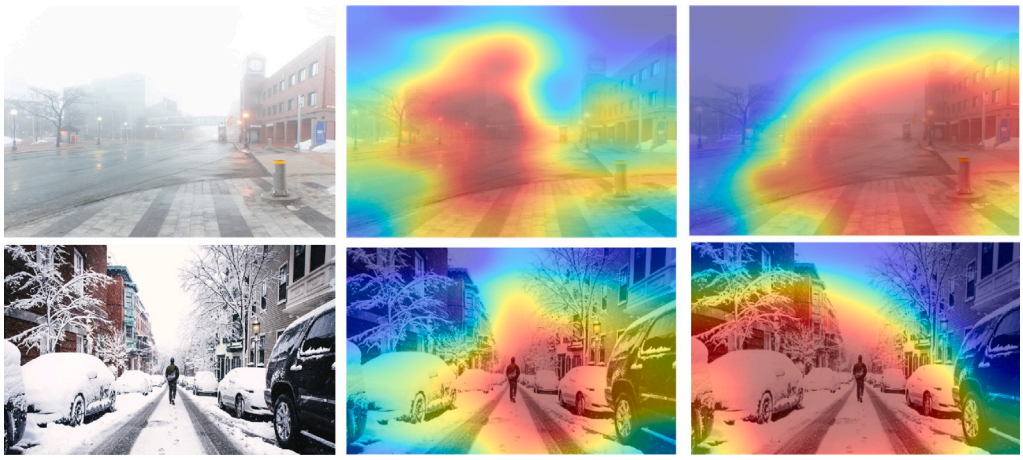


Fig. 6. Two samples (left) and the corresponding heatmap depicting the class activation maps for each classification branch, namely weather (center) and ground surface (right). The heatmap may help to interpret the network decision by highlighting the regions that mostly contribute to each network prediction.

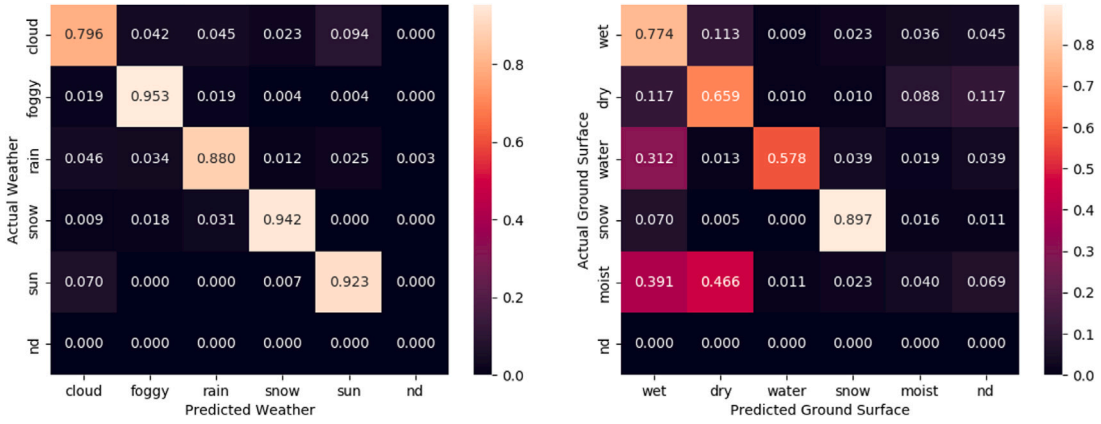


Fig. 7. Confusion matrices of a multi-task neural network trained with the masked asymmetric loss using uniform γ^- weights on the considered classes for weather classification and ground surface recognition. Not considering the distribution of the training in the choice of the weights, the variance of the classification accuracy over the categories increases with respect to the performance achieved by the proposed solution.

surface classification. The use of a spatial-wise and channel-wise attention mechanism such as CBAM, which proved more effective than HAM in the ablation study, further helped the multi-task network to generalize across both the tasks; the selection of diversified regions of interest using the two task-specific attention branches allows the neural network to focus on the most relevant regions of the image for the task of interest. The obtained results are not only encouraging in terms of classification accuracy, but also for the saving of resources and processing time with respect to the sequential execution of the two single-task neural networks. Therefore, the proposed solution represents a tool that can be used in real applications for quick intervention in case of dangerous or non-optimal weather or ground surface.

The dataset collected and annotated for the training of the multi-task neural network proved to be sufficiently representative of the problem, considering that the performance of the trained system was verified on a test set composed of totally different images. Having demonstrated the effectiveness of this procedure, it may be possible to extend the dataset by exploiting the presence of numerous public live webcams around the world. The scenarios may be chosen by considering the imbalance among the classes, in order to have a more balanced distribution of the samples. The annotation of these new images can be carried out with a reliable but expensive manual procedure, or with a semi-automatic procedure based on data imputation techniques. In particular, the latter can be done by using already trained networks as annotators, since the achieved accuracy, especially on weather recognition, allows to speed up the process; of course, these labels obtained automatically should be then verified manually.

The use of a masked loss allowed to exploit all the samples in the dataset, even those partially annotated; considering that there were almost 27,000 samples not annotated for weather classification and over 12,000 for ground surface classification, the advantage of using so many samples to improve the individual classification heads is not negligible. Having a very accurate neural network available, even if possibly more complex and computationally expensive, an alternative may be to obtain the missing labels through knowledge distillation; in this way, it would be possible to train a student multi-task neural network starting from a dataset with complete labels produced by expert masters.

The proposed learning procedure, based on asymmetric loss, proved to be effective even with an unbalanced distribution of data within the training set; the experimental results demonstrated the validity of this choice with respect to a cross entropy loss. As it is evident from Fig. 7, also the choice of the γ^- weights plays a fundamental role; in fact, by setting uniform values equal to 1, the variance of the accuracy on the various classes increases, penalizing the categories with less samples in the dataset (e.g. snow and moist). Of course, by using a dataset with a more balanced distribution of samples among the classes the cross entropy loss may be the best solution.

By following these research directions, the functionalities described in the paper can be effectively used in modern smart roads to provide real-time feedback on weather and road conditions, enabling adjustments in streetlight intensity or timely alerts to authorities, identifying and localizing hazardous weather or asphalt conditions for drivers and

improving agriculture production by prompt interventions of farmers or autonomous robots for watering, irrigation, and protecting plants from adverse weather and frost.

6. Conclusions

In this paper we described a multi-task neural network for real-time recognition of weather and ground surface conditions. This advanced artificial vision technique has proven to be an effective tool for the immediate notification of anomalous meteorological situations to guarantee preventive intervention to safeguard agricultural heritage and smart roads, as well as the safety of people. The proposed solution achieved higher accuracy on the two tasks of interest compared to standard single-task neural networks and existing methods, while guaranteeing significant savings in processing time and computing resources. The procedure for collecting and annotating the dataset to train the multi-task neural network, as well as the learning process based on attention mechanisms, masked asymmetric loss and GradNorm, have demonstrated their effectiveness in the experimental analysis and constitute a first step towards creating even more effective and efficient in a field in which artificial vision has great margins for development. Future developments in the collection of increasingly representative datasets and more effective and efficient architectures can pave the way for the creation of video surveillance systems with smart cameras, installed in rural and urban areas, for the protection of agricultural heritage and road safety.

CRediT authorship contribution statement

Diego Gragnaniello: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization. **Antonio Greco:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization. **Carlo Sansone:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Conceptualization. **Bruno Vento:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Data curation, Conceptualization.

Declaration of competing interest

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

Acknowledgment

We acknowledge financial support from: PNRR MUR project PE0000013-FAIR.

Data availability

Data will be made available on request.

References

Abdelraouf, A., Abdel-Aty, M., Wu, Y., 2022. Using vision transformers for spatial-context-aware rain and road surface condition detection on freeways. IEEE Trans. Intell. Transp. Syst. 23 (10), 18546–18556. http://dx.doi.org/10.1109/TITS.2022.3150715.

Al-Haija, Q.A., Smadi, M.A., Zein-Sabatto, S., 2020. Multi-class weather classification using ResNet-18 CNN for autonomous IoT and CPS applications. In: International Conference on Computational Science and Computational Intelligence. CSCI, IEEE, pp. 1586–1591.

Almazan, E.J., Qian, Y., Elder, J.H., 2016. Road segmentation for classification of road weather conditions. In: ECCV 2016 Workshop. Springer, pp. 96–108.

Carrillo, J., Crowley, M., Pan, G., Fu, L., 2020. Design of efficient deep learning models for determining road surface condition from roadside camera images and weather data. arXiv preprint arXiv:2009.10282.

Chen, Z., Badrinarayanan, V., Lee, C.-Y., Rabinovich, A., 2018. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: International Conference on Machine Learning. PMLR, pp. 794–803.

Chen, S., Shu, T., Zhao, H., Tang, Y.Y., 2023. MASK-CNN-transformer for real-time multi-label weather recognition. Knowl.-Based Syst. 278, 110881.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection. In: International Conference on Computer Vision and Pattern Recognition. Vol. 1, IEEE, pp. 886–893.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: A large-scale hierarchical image database. In: International Conference on Computer Vision and Pattern Recognition. IEEE, pp. 248–255.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

Elhoseiny, M., Huang, S., Elgammal, A., 2015. Weather classification with deep convolutional neural networks. In: 2015 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 3349–3353.

Foggia, P., Greco, A., Saggese, A., Vento, M., 2023. Multi-task learning on the edge for effective gender, age, ethnicity and emotion recognition. Eng. Appl. Artif. Intell. 118, 105651.

Gao, L., 2019. Five class weather image dataset. http://dx.doi.org/10.21227/ffbx-0p16.

Gbeminiyi, A., 2018. Multi-class weather dataset for image classification. Mendeley Data 6, 15–23.

Grabowski, D., 2020. Road CCTV images with associated weather data. http://dx.doi.org/10.7910/DVN/SV9N9F.

Gragnaniello, D., Greco, A., Sansone, C., Vento, B., 2024a. Fire and smoke detection from videos: A literature review under a novel taxonomy. Expert Syst. Appl. 124783.

Gragnaniello, D., Greco, A., Sansone, C., Vento, B., 2024b. Smart visual sensors for real time weather and ground conditions recognition for agricultural robotics. In: International Conference on Automation Science and Engineering. CASE, IEEE, p. to appear.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: International Conference on Computer Vision and Pattern Recognition. pp. 4700–4708.

Jun, S., Jang, H., Kim, S., Lee, J.-S., Jung, D., 2024. A review of ground camera-based computer vision techniques for flood management. Comput. Concrete 33 (4), 425.

Kanakis, M., Bruggemann, D., Saha, S., Georgoulis, S., Obukhov, A., Van Gool, L., 2020. Reparameterizing convolutions for incremental multi-task learning without task interference. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16. Springer, pp. 689–707.

Kang, L.-W., Chou, K.-L., Fu, R.-H., 2018. Deep learning-based weather image recognition. In: International Symposium on Computer, Consumer and Control (IS3C). IEEE, pp. 384–387.

Khan, M.N., Ahmed, M.M., 2022. Weather and surface condition detection based on road-side webcams: Application of pre-trained convolutional neural network. Int. J. Transp. Sci. Technol. 11 (3), 468–483.

Lagorio, A., Grosso, E., Tistarelli, M., 2008. Automatic detection of adverse weather conditions in traffic scenes. In: International Conference on Advanced Video and Signal Based Surveillance. IEEE, pp. 273–279.

Li, G., Fang, Q., Zha, L., Gao, X., Zheng, N., 2022. HAM: Hybrid attention module in deep convolutional neural networks for image classification. Pattern Recognit. 129, 108785.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In: IEEE/CVF International Conference on Computer Vision. pp. 10012–10022.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022. A convnet for the 2020s. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11976–11986.

Lu, C., Lin, D., Jia, J., Tang, C.-K., 2017. Two-class weather classification. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2510–2524. http://dx.doi.org/10.1109/TPAMI.2016.2640295.

Mittal, S., Sangwan, O.P., 2023. Classifying weather images using deep neural networks for large scale datasets. Int. J. Adv. Comput. Sci. Appl. 14 (1).

Nolte, M., Kister, N., Maurer, M., 2018. Assessment of deep convolutional neural networks for road surface classification. In: International Conference on Intelligent Transportation Systems. ITSC, IEEE, pp. 381–386.

Onesimu, J.A., Kadam, A., Sagayam, K.M., Elngar, A.A., 2021. Internet of things based intelligent accident avoidance system for adverse weather and road conditions. J. Reliable Intell. Environ. 1–15.

Qian, Y., Almazan, E.J., Elder, J.H., 2016. Evaluating features and classifiers for road weather condition analysis. In: IEEE International Conference on Image Processing. ICIP, IEEE, pp. 4403–4407.

Ridnik, T., Ben-Baruch, E., Zamir, N., Noy, A., Friedman, I., Protter, M., Zelnik-Manor, L., 2021. Asymmetric loss for multi-label classification. In: IEEE/CVF International Conference on Computer Vision. pp. 82–91.

Samo, M., Mafeni Mase, J.M., Figueredo, G., 2023. Deep learning with attention mechanisms for road weather detection. Sensors 23 (2), 798.

Sandnes, F.E., 2012. Generating weather reports using automatically classified webcam images. Int. J. Comput. Consumer Control 1 (1), 1–7.
Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In: International Conference on Machine Learning. PMLR, pp. 6105–6114.
Woo, S., Park, J., Lee, J.-Y., Kweon, I.S., 2018. Cbam: Convolutional block attention module. In: European Conference on Computer Vision. ECCV, pp. 3–19.
Xie, K., Huang, L., Zhang, W., Qin, Q., Lyu, L., 2022. A CNN-based multi-task framework for weather recognition with multi-scale weather cues. Expert Syst. Appl. 198, 116689.
Zhang, Z., Ma, H., Fu, H., Zhang, C., 2016. Scene-free multi-class weather classification on single images. Neurocomputing 207, 365–373.
Zhao, T., He, J., Lv, J., Min, D., Wei, Y., 2023. A comprehensive implementation of road surface classification for vehicle driving assistance: Dataset, models, and deployment. IEEE Trans. Intell. Transp. Syst..
Zhao, B., Hua, L., Li, X., Lu, X., Wang, Z., 2019. Weather recognition via classification labels and weather-cue maps. Pattern Recognit. 95, 272–284.