Pre-computation of image features for the classification of dynamic properties in breaking waves

Ryan Smith, Frédéric Dias, Gabriele Facciolo, Thomas Brendan Murphy

Letter

Pre-Computation of Image Features for the Classification of Dynamic Properties in Waves

Ryan Smith [1], Frédéric Dias [1,2], Gabriele Facciolo[2] and Thomas Brendan Murphy [1]

[1] School of Mathematics and Statistics, University College Dublin, Ireland
[2] Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190 Gif-sur-Yvette, France
* Correspondence: frederic.dias@ucd.ie

Version November 17, 2023 submitted to Preprints

Abstract: The use of convolutional neural networks (CNNs) in image classification has become the standard method of approaching computer vision problems. Here we apply pre-trained networks to classify images of non-breaking, plunging and spilling breaking waves. The CNNs are used as basic feature extractors and a classifier is then trained on top of these networks. The dynamic nature of breaking waves is exploited by using image sequences to gain extra information and improve the classification results. We also see improved classification performance by using pre-computed image features such as the optical flow between image pairs to create new models in combination with infra-red images. The inclusion of the dynamic information improves the classification between breaking wave classes. We also provide corrections to the methodology from the article from which the data originates to achieve a more accurate assessment of performance.

Keywords: Breaking waves; Optical flow; Convolutional Neural Networks; Image Classification

1. Introduction

Large ocean waves carry huge amounts of energy which can be dissipated through a process known as breaking. One example of breaking is when an overturning of the crest causes a collapse and a breakdown to turbulence. The breaking process releases large amounts of energy in the dissipation of the kinetic energy of the wave through this turbulence. Understanding the breaking process, exchange of gases and energy in waves at various scales is of great importance to improving models of ocean-atmosphere interactions such as weather and climate models [1,2]. The dissipation of large amounts of energy through turbulent breaking is also of interest in coastal engineering applications where waves may slam into cliffs or man-made coastal structures.

The dissipation of energy at small length-scales creates difficulty in modelling breaking waves, as they interact with the air above the surface in the breaking process, creating large and small bubbles through a turbulent air entrainment [3,4]. Breaking waves being turbulent, it is a complex two-phase process at the free surface where by overturning waves generate bubbles and jets of water plunging back into the wave create vortices which increase the mixing further. Current numerical methods make great simplifications/assumptions about the fluid flow at which point the details in the breaking process may be lost [5] and what happens after breaking is largely unresearched as numerical methods are unable to simulate the process on large scales.

Ocean waves are difficult to recreate in the laboratory due to the different salinity, temperature and thus density gradients throughout the fluid. The waves are usually generated mechanically and do not experience the same breaking or spraying as ocean waves generated by wind. It is desirable then to study breaking waves so that it is feasible to measure turbulent quantities and determine characteristics of these waves in a real world setting. Examples are the breaking threshold, a criterion that distinguishes breaking waves from non-breaking waves and the spreading of breaking through the wave.

To overcome these challenges new methods are being pursued to extract dense or detailed information from ocean waves in real time using video data. A large video database of waves allows

for a thorough analysis of breaking waves using image processing and modern computer vision techniques. The collected image data can also be used in the training of deep learning models for classification [6], clustering, segmentation and prediction of wave characteristics [7]. Such models can be used in further processing the image data collected to automatically detect, track and estimate quantitative wave properties.

We have applied image processing algorithms to breaking wave image data to probe what extra information can be gained from this approach of analysing fluid flows and in particular breaking waves. The main technique used on the dataset is optical flow [8,9], which calculates the displacement of pixels between a pair of images. One of the difficulties is the discontinuous nature of breaking waves and the fact that these algorithms have been developed to deal with typically more rigid objects and motions. Thus it may prove necessary to use images from video at high resolution and high frequency to be able to resolve the details in the motions.

2. Related Works

Advances in the field of machine learning (ML) and increased processing power of computers over the last two decades have seen ML incorporated into almost every quantitative research field. Fluid dynamics and oceanography have been no different, with applications from the simulation of fluids [10] to flow optimisation and control [11]. Image processing and computer vision techniques have also shown great promise in oceanographic research where both classical algorithms and deep learning approaches have been implemented to process large amounts of image data to produce three dimensional depth maps [12–14], wave tracking [15], classification and segmentation of waves [16,17] and wave height estimations [18,19].

Infra-red provides somewhat of an advantage for the study of breaking waves when using images due to small temperature fluctuations which are recorded by the infrared camera at the breaking region [20]. These temperature fluctuations manifest as streaky patterns on the back of plunging waves [21] and can be detected with the feature extraction in modern convolutional neural networks. The infra-red cameras are also less sensitive to the reflection of light from the sun on the sea surface than traditional visible wavelength cameras. Using infra-red cameras does however come with limitations in image resolution and observation distance.

Optical flow in fluids research has also been used to analyse the motion of fluids in laboratory wave tanks [5,22] and to improve the results from Particle Image Velocimetry (PIV) [23,24]. Attempts to constrain the variational equations of optical flow with fluid mechanics equations has shown increased accuracy in some cases [25–27].

Video classification algorithms where both raw frames and optical flows are combined to help with the classification task have also been explored [28] and shown to give increased performance over the same algorithms without the additional optical flow information. Although the cited article uses a more complex Long-Short-Term-Memory (LSTM) network architecture, which was not implemented for our experiments as we use only two frames compared to a longer video sequence, they explore the usefulness of additional optical flow information. However it is noted that optical flow on its own in a case of noisy video frames can show degraded classification performance in comparison to just raw image frames.

3. Materials and Methods

3.1. Data collection

The data used for this project is from Buscombe and Carini [6] in which they used a multitude of popular pre-trained CNNs as basic feature extractors for infra-red (IR) images of breaking waves in order to classify them. Details of the data acquisition can be found within the cited paper. The dataset consists of 9996 images split among three different wave classes: nonbreaking, plunging and spilling waves. The IR images were taken at a resolution of 640 by 480 pixels and are downsampled

to a resolution of 299 by 299 for the CNN feature extractions. The dataset is highly imbalanced and contains relatively few examples of the plunging breakers: with 9996 total images, 208 are plunge, 2354 spilling and 7434 non-breaking wave images. Consecutive sample frames from each one of the wave classes are shown in Fig. 1.

As documented in Buscombe and Carini [6], the images were captured using a thermal infra-red camera in November of 2016 at a US Army Corps of Engineers Field Research Facility in Duck, North Carolina, USA. The camera was angled at 45° to the sea surface and mounted on a pier at the research facility. The images were sampled continuously at 10 frames per second while a Light Detection and Ranging device (LIDAR) measured the sea surface elevation in the same field of view. The wave height, as measured by the LIDAR, varied between 0 and 5.94 meters during the 10.5 hour acquisition period. The sequences vary in length from 7 frames to over 700 frames in a sequence (shorter sequences typically belonging to the plunge and spill breaking waves, while nonbreaking sequences are much longer). The distribution of pixel intensities was used to determine if the image contained a breaking wave. The type of breaking wave detected in these images was then manually classified by examining the patterns in front of and behind the breaking region. We made several changes to the dataset after finding errors in the manual classification which are described in subsection 3.2. Further details of camera and LIDAR specifications are available in the cited paper.

Our proposal is to achieve better results on the classification task by incorporating some of the dynamical information from the waves using the optical flow methods described in the following sections. In the original paper, the authors claim to have a high precision and recall on each of the classes (after augmentation F_1 scores ≈ 92 for all classes) but in a further analysis we found flaws with their dataset and model. The metrics used to assess the models were significantly degraded after correction of the dataset problems described in subsection 3.2. Thus, we aim to improve upon the corrected results by inclusion of the dynamical information.

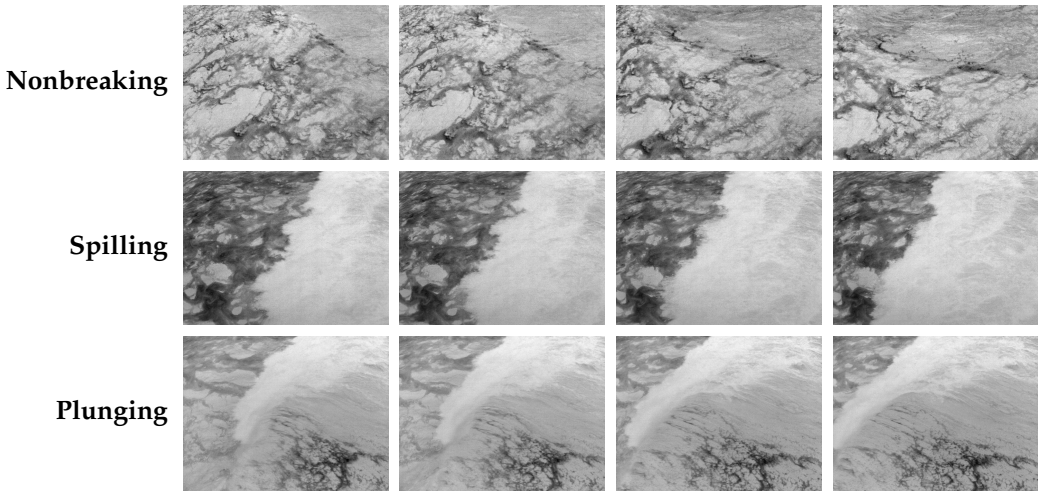

Figure 1. Samples of wave dynamics for each of the given classes. The image sequences span a time-frame of 1.0, 0.4 and 0.4 seconds for the non-breaking, spilling and plunging sequences respectively. The wave movement is difficult to observe in the non-breaking case. However the breaking waves (spilling and plunging) show turbulent white-water effects which are picked up well by the infra-red (IR) camera. These white-water patterns contain distinctive identifying markers for the type of breaking that occurs. In the spilling case the breaking occurs more slowly and the white-water spreads over the back face of the wave. To contrast this in the plunging wave there is more structure to the back of the wave as the breaking happens more suddenly and crashes in front of the wave.

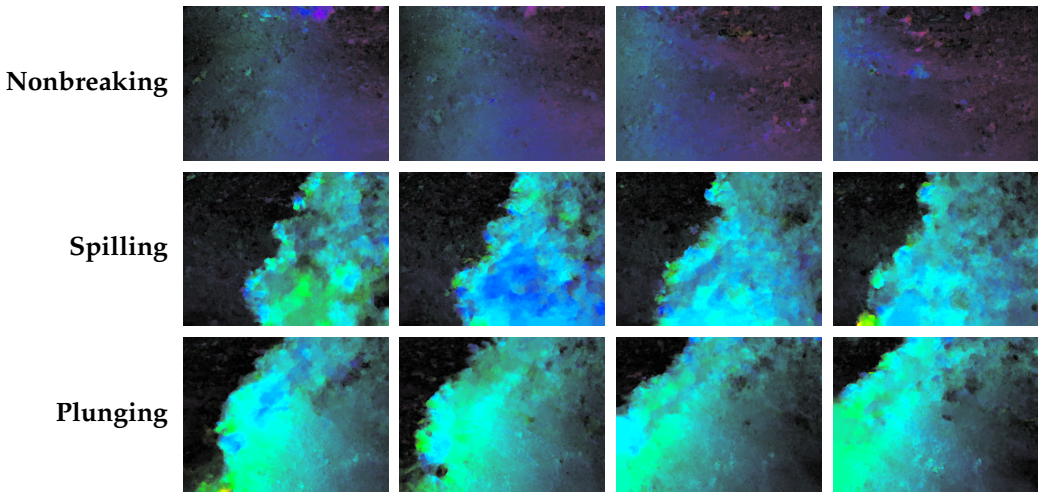

Figure 2. Samples of calculated optical flows using TV-L[1] for each of the given classes. The optical flows are calculated using the sample IR images in Figure 1 and their respective next frame. The difference between the non-breaking and breaking classes is clear while the differences between spilling and plunging classes are more subtle but the extended "white water" region can be seen in the spilling optical flows.

3.2. Dataset Corrections

During the investigation several issues were identified with the dataset that was originally proposed to be used. Firstly in going through the original authors code, it was noted that the dataset splitting into training and testing sets was randomised. As the data was sequential (video images at 10 frames per second) it essentially meant that the authors' model was training on images that were one tenth of a second away from the testing images and in most cases the model saw images either side of the test image. Thus to sort the dataset correctly we separated the images into discrete waves based on the filename (i.e. group images xxx110 xxx120 and xxx130 into wave 1 for the class they belong to).

The grouping of the images into waves revealed another problem with the dataset: misclassified images. The grouping method described above revealed some waves in the classes contained only one or two images, but a file search for the images that would surround these revealed images of the same wave put into a different class (e.g. found spill/xx30, spill/xx50 but plunge/xx40). This is a clear mislabelling of the dataset as the waves do not change class and back again within 0.2 seconds. Each of these corrections was manually verified.

Once the data was sorted correctly, we tested the original authors best model and found it was significantly less accurate and the original logistic regression model did not perform well even on training data. One of the proposed reasons for this under-performance is the large class imbalance in the dataset and the fact that there are very few samples of the plunging breaking waves. Thus, the model is unable to find a distinguishing feature between the spilling waves and the plunging waves before over-fitting to the training set. It is cautioned to not randomly split sequential data like this which is highly dependent on the previous observation when splitting into training and testing datasets. This will give false results in regard to the models ability to generalise as the testing set contains near identical examples to the training set.

3.3. Optical flow

Motion analysis in image processing can be reduced to three main groups of problems: motion detection, object tracking and location and the derivation of 3D properties from 2D projections acquired at different points in time [29]. In this section we are concerned with tracking the motions of individual pixels in a pair of images. In this way we calculate the displacement of pixels from the first image to the second; this is optical flow.

The optical flow between a pair of successive images is the apparent motion of the objects appearing in the two frames [29]. Thus the optical flow is a vector field of the displacements of pixels and provides a mapping from the first image to the second image. Errors in this mapping can be caused by the following features: occlusion where the movement of objects blocks other objects in the second image, a lack of texture (high gradients in the pixel intensity) on the objects, or sudden brightness shifts such that tracking the motion becomes difficult.

A variety of techniques for calculating the optical flow field exist including classical methods and methods using convolutional neural networks (CNN) for the estimation of optical flow. Classical methods include variational techniques like the Horn-Schunck optical flow [8] which minimises an energy functional over the images, yielding a dense (for each pixel) vector field of displacements. These techniques are based on the assumption that displacements are only one or two pixels in size, thus pyramid or multi-scale approaches are implemented which allow the detection of larger motions. The multi-scale method downsamples the image to calculate a rough optical flow to be used as a base approximation for each increase in resolution back to the original resolution. This is often referred to as a coarse-to-fine optical flow estimation.

The CNN based methods are instead trained to estimate the motion fields in a more "black box" manner by providing many example images and their ground truth optical flows. A CNN model then learns to detect the textures within the images, and the change in location of these textures between the images thus producing an optical flow field. The ability to generalise is then dependent on the variety of textures and objects in motion in the images used for training the CNN. Datasets used for this training create an artificial optical flow by overlaying objects in the images and moving them from one frame to the next, in this way they know the "ground truth" optical flow.

We have selected two optical flow (OF) methods for use in this study; a classical method TVL^1 [9, 30–32] and a CNN based method SPyNet [33,34]. The details of the OF algorithms can be found in the respective articles along with working implementations. Fig. 2 shows the optical flow (computed with TVL^1) corresponding to the sample frames shown in Fig. 1.

3.4. Training and classification metrics

For the feature extraction, we use different pre-trained image classification models from the Tensorflow Python library. The top (fully connected) layer of these networks was removed so that a vector of features was outputted. This feature vector is then saved for each image or image pair. For comparison with the original paper [6], we then trained the same logistic regression (LR) model on these feature vectors. A new fully connected layer was found to give no significant improvement over the LR model although more samples of the plunging class or more aggressive data augmentation could boost the performance of the neural network but this was not pursued here. The images were downsampled to the appropriate size for each respective CNN architecture (based on what they had originally been trained on) for optimal feature extraction. We trained several different models of which some take two or three images as the inputs. For the multi-image inputs we process each image with the CNN and combine the feature vectors together before training the classification model. Stacking the inputs on top of each other and processing these with the CNNs was found to produce inferior results. This result is likely because the channels in these pre-trained CNNs corresponded to the red, green and blue features but extracting all those channel features on each grayscale image gave more useful information. We trained an IR, OF, IR+IR and IR+OF models, where IR stands for infra-red and OF stands for optical flow image inputs. The addition represents the concatenation of the feature vectors with IR+IR being two consecutive infra-red images.

The metrics used for the evaluation of the classification predictions are described below, with the inclusion of two more metrics than in the Buscombe and Carini paper. The models were assessed using five different metrics on the classes: Precision (Pr), Recall (Re), F1 score (F1), Informedness [35] (In) and the Brier score [36] (Bs) and each is defined in terms of the true and false positive (TP, FP) and true and false negatives (TN, FN) as follows:

$$Pr = \frac{TP}{TP + FP}, \quad Re = \frac{TP}{TP + FN}, \quad F1 = 2\frac{Pr \times Re}{Pr + Re}$$

$$In = \frac{TP}{TP + FN} + \frac{TN}{TN + FP} - 1, \quad Bs = \frac{1}{N}\sum_{t}^{N}\sum_{i}^{R}(p_{ti} - l_{ti})^2,$$

where in the Bs the sums are over all N samples and all R classes, p_{ti} represents the predicted probability for sample t and class i and l_{ti} is then a vector with a 1 in the position indicating the true label. It thus measures the mean square difference between the predicted probability and the actual labels and thus a lower Bs is better. The informedness score estimates the probability of making an "informed decision": a score of 1 indicates a perfect classifier while a score of 0 indicates random decisions.

4. Results

4.1. Classification

After fixing the problems with the original dataset (see subsection 3.2), we use the original authors methods as a new baseline measure for our results. The best model from the [6] paper is a logistic regression model fit to features extracted by a pre-trained MobileNet_V2 [37]. We found the MobileNet_V2 CNN to perform much worse than using the features extracted from the Xception [38] pre-trained CNN. All model results below are from features extracted by the Xception network and a logistic regression fit.

Table 1. Metrics for tested models. For each metric (column) the best score for each class is indicated by the bold text. Higher scores are better for all metrics except for the Brier score where a score of 0 indicates perfect predictions and confidence. Results for models using augmented data are in parenthesis.

Model	**Class**	Precision	Recall	F_1 score	Informedness	Brier score
IR	Non-breaking	0.97 (0.96)	0.94 (0.93)	0.95 (0.95)	0.87	0.09
	Plunge	0.33 (0.30)	0.42 (0.47)	0.37 (0.37)	0.41	0.97
	Spill	0.80 (0.79)	0.85 (0.83)	0.82 (0.81)	0.79	0.22
TV-L[1]	Non-breaking	**0.98 (0.98)**	0.96 (0.96)	0.97 (0.97)	**0.90**	0.06
	Plunge	0.32 (0.30)	**0.59** (0.57)	**0.42** (0.39)	**0.58**	**0.70**
	Spill	0.85 (0.85)	0.86 (0.85)	0.85 (0.85)	0.81	0.22
SPyNet	Non-breaking	0.97 (0.97)	0.96 (0.96)	0.96 (0.96)	0.87	0.06
	Plunge	0.15 (0.14)	0.46 (0.46)	0.22 (0.22)	0.43	0.92
	Spill	0.84 (0.84)	0.76 (0.76)	0.80 (0.80)	0.72	0.33
IR + IR	Non-breaking	**0.98** (0.97)	0.96 (0.95)	0.97 (0.96)	0.89	0.08
	Plunge	0.39 (0.34)	0.43 (0.49)	0.41 (0.40)	0.42	1.01
	Spill	0.84 (0.82)	0.89 (0.87)	0.86 (0.84)	0.83	0.24
IR + TV-L[1]	Non-breaking	**0.98 (0.98)**	**0.97 (0.97)**	**0.98 (0.98)**	**0.90**	**0.05**
	Plunge	**0.43** (0.36)	0.41 (0.46)	**0.42** (0.40)	0.48	0.91
	Spill	**0.88 (0.88)**	**0.92** (0.91)	**0.90** (0.89)	**0.86**	**0.12**

From Table 1 we can see that the optical flow of the images can be used in the classification task and it provides results comparable to using a single image. However the best results came from using the two IR images or a combination of the IR images and the optical flow. The extra information from the use of two images does not affect the nonbreaking classifications significantly, but it does have a positive impact on the spill and plunge classifications. In our experiments we also observed that the image augmentation implemented (rotation, zoom, crop) does not help when using the optical flow and in some cases it decreases performance. The SPyNet model shows difficulty in extracting the features to differentiate the two breaking types (plunge and spill). The CNN's output for calculating optical flow was seen to produce inaccurate and overly smooth flows on these images and thus

performs slightly worse than the TV-L^1 model. The probabilities of each class obtained from the models are visualised in Fig. 4.

We also tabulate the errors from each of these metrics to make it more clear where exactly the best improvements in our models are. In Table 2 we give these improvements as percentages relative to the baseline IR model. It is clear that most models have performed significantly better in most metrics when compared to the single IR input model. The IR+TV-L^1 model reports the majority of largest improvements (indicted by the **bold** text).

Table 2. The table below gives first the Infrared errors as a reference in section (a). The rest of the rows within section (b) contain the relative % improvement over the simple IR input model. Each error (except for Brier score) is calculated as $1 - score$, where the score is the value reported in Table 1. For the Brier score, since a lower score is best, we leave it as is. The percentages reported are the improvements in respective errors for the models, so a high positive % indicates a large improvement in the metric (reduced error). The best improvements are once again highlighted in bold text.

(a)

Errors						
Model	**Class**	Precision	Recall	F_1 score	Informedness	Brier score
IR	Non-breaking	0.03	0.06	0.05	0.13	0.09
	Plunge	0.77	0.58	0.63	0.59	0.97
	Spill	0.20	0.15	0.18	0.21	0.22

(b)

Relative Improvements over IR Model Errors						
TV-L^1	Non-breaking	**+33.33%**	+33.33%	+40.0%	**+23.08%**	+33.33%
	Plunge	-1.49%	**+29.31%**	**+7.94%**	**+28.81%**	+27.84%
	Spill	+25.0%	+6.67%	+16.67%	+9.52%	0.0%
SPyNet	Non-breaking	0.0%	+33.33%	+20.0%	0.0%	33.33%
	Plunge	-26.87%	+6.9%	-23.81%	+3.39%	+5.15%
	Spill	+20.0%	-60.0%	-11.11%	-33.33%	-50.00%
IR+IR	Non-breaking	**+33.33%**	+33.33%	+40.0%	+15.38%	+11.11%
	Plunge	+8.96%	+1.72%	+6.35%	+1.69%	-4.12%
	Spill	+20.0%	+26.67%	+22.22%	+19.05%	-9.09%
IR+TV-L^1	Non-breaking	**+33.33%**	**+50.0%**	**+60.0%**	**+23.08%**	**+44.44%**
	Plunge	**+14.93%**	-1.72%	**+7.94%**	+11.86%	+6.19%
	Spill	**+40.0%**	**+46.67%**	**+44.44%**	+33.33%	**+45.45%**

4.2. Misclassifications

In this section we present the analysis of some misclassifications and confusion matrices in Fig. 3 of the models on testing (unseen) data both with and without augmentation applied. This allows for quick visual identification of misclassifications by looking at the off-diagonal terms in the confusion matrices. For the confusion matrices, each row corresponds to a true label and each column corresponds to the predicted label. Higher performance corresponds to a darker main diagonal in the matrix. We observe that all models perform well on the non-breaking waves, with IR and the SPyNet OF having the largest errors on spilling waves (approximately 19 and 10 percent of spilling waves classified as non-breaking respectively). The TV-L^1 OF and the combined IR+OF (TV-L^1) are the best performing models at separating plunging waves from spilling waves. The majority of misclassifications occur as a wave is just entering or exiting the frame when it is reasonable to be incorrect.

5. Discussion

We have explored the use of optical flow features for the classification of breaking wave images. This was tested on an image classification task to gain improved classification of sequences of IR images of breaking waves compared to only using a single IR image. After an initial exploration of the original dataset, the original analysis was found to be flawed. A large class imbalance exists,

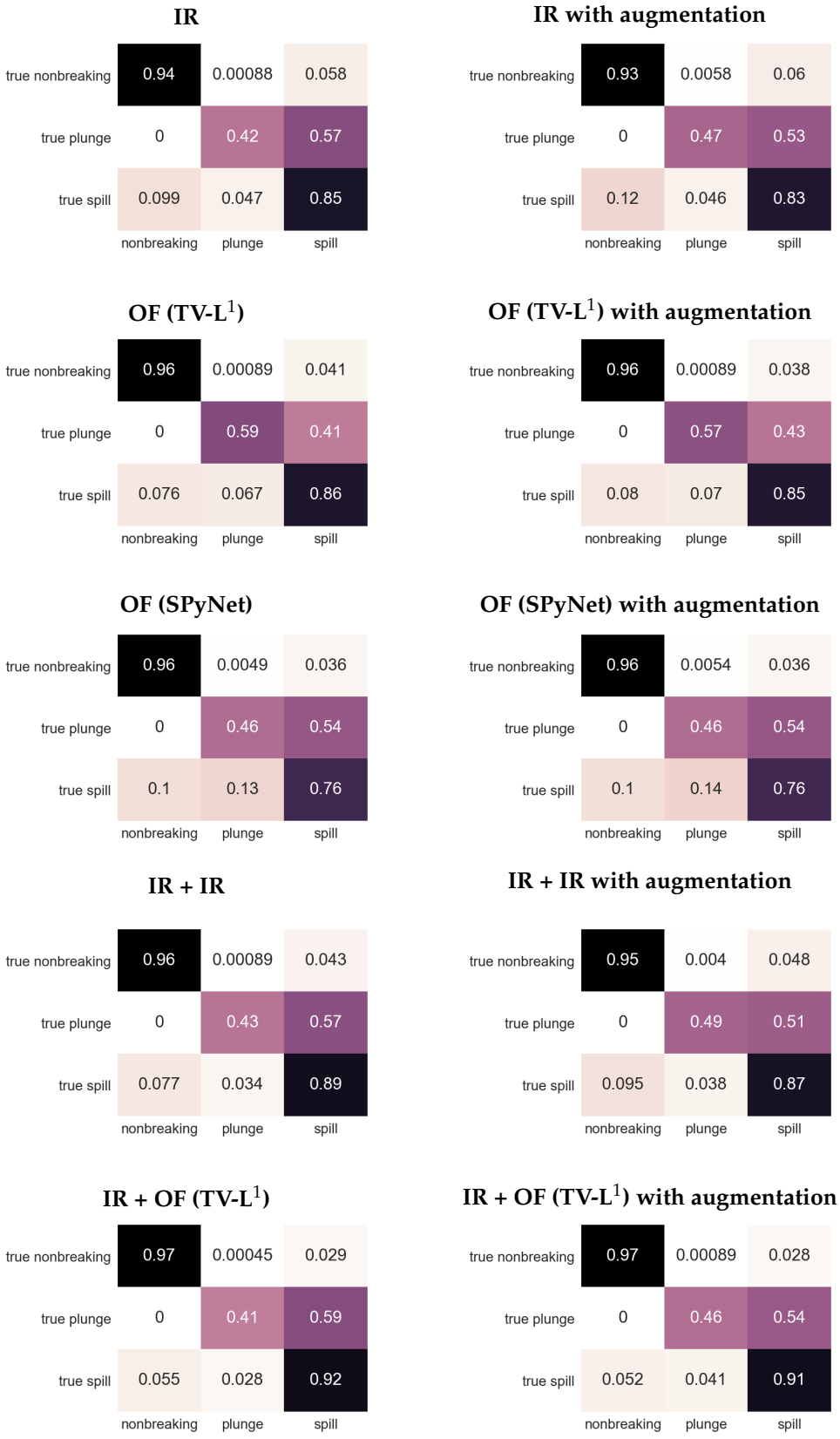

Figure 3. Confusion matrices for each of the tested models. The true class is in each row and the predicted class in each column.

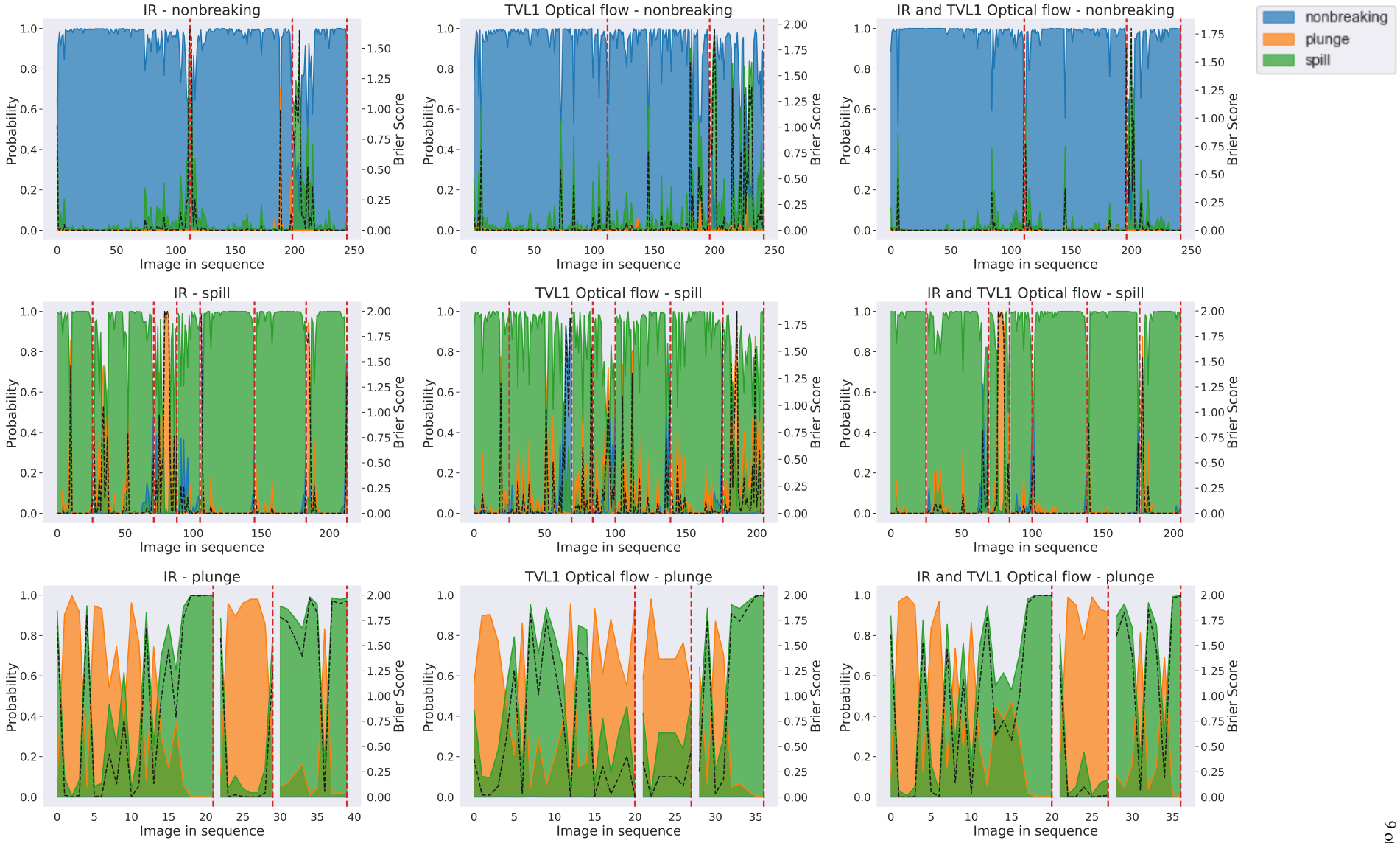

Figure 4. The evolution of probabilities for each class from some of the tested models. The horizontal axis corresponds to the image in the sequences. We see that the gain from the optical flow method in the classification of plunge images comes at the cost of uncertainty in the spill cases. A Brier score of 0 indicates a correct classification and total confidence in the classification i.e. the probability of that class is 1. Sudden spikes or changes in predictions are mostly from start and ending of image sequences where waves are only starting to enter the frame or almost entirely out of frame.

which made it difficult to train and test a robust classifier, and initial results were compromised. After correction of the data, comparisons were drawn between the baseline IR model which was claimed as best by the original authors and several other models which took advantage of the dynamical nature of the waves by performing a feature extraction on sequential images. It was found that a novel combination of the IR and optical flow information could produce slightly better results and the inclusion of augmented images provides additional gains to the IR models, but acts to somewhat deteriorate the performance of the optical flow models.

Our experiments have focused on the inclusion of temporal features in the data used to train the models. This gives a significant reduction in errors when compared to the baseline, single infra-red image input, model. In contrast to Buscombe and Carini [6], who have quantified the performance of different pre-trained network feature extractions, our focus was thus on the data input to the networks.

As opposed to the results in table 1 of [6], we detected and removed a selection bias within the data split caused by the high correlation of the training and testing data when randomly selected. Our results give a baseline which corresponds to Buscombe and Carini's results for the Xception network when the selection bias is removed and all the models, with the exception of the SPyNet based optical flow, have shown significant improvements over the baseline model. Further our results show that our application of optical flow as input to a pre-trained CNN has reduced errors on the classification task but similar to Ng et al. [28] noisy optical flows lead to a deteriorated performance as with the SPyNet results.

An analysis of the misclassified cases of the models shows that most misclassifications occur at the beginning or end of the image sequences where the type of wave is unclear. The gains on identifying both the plunging and spilling waves from using the optical flow can be achieved also by using the two IR images and applying the feature extraction on these. Further improvements may be possible with more regularisation or more aggressive data augmentation, in combination with more complex models, for better performance on the plunging samples. The dataset was deemed insufficient to train more complex neural networks or fine-tuning without over-fitting and adapting of the SPyNet optical flow network was abandoned after it failed to produce accurate optical flows on the images. The inclusion of Optical Flow into the features for the classification task improved the results on the plunging wave category, and a combination of optical flow and IR images gave the best results on the spilling waves.

In future work, a higher quality dataset in terms of clearer and higher resolution images, but also in terms of a better balance of the breaking wave classes, will make the classification task and training of more complex models easier. Distinctive patterns in wave breaking can be identified but capturing the temporal evolution of wave properties through a sequence of images remains a challenge. Investigation of how breaking evolves over time and an understanding of how it affects the free surface is crucial to provide accurate parameterizations for numerical forecasting systems.

6. Conclusions

The introduction of optical flow to incorporate temporal features leads to significant increase in a models ability to distinguish between plunging and spilling waves. The novel application of optical flow (TV-L^1 and SPyNet) and sequential images to breaking wave to infer temporal features yields significant performance gains as evident from both Table 2 and in Fig. 3 where errors are reduced by up to 60%. Moreover, our use of sequential images has shown that improved performance was achieved in nearly every metric for each model tested versus a baseline result obtained using only a single infra-red image. The addition of image augmentation to boost the number of image samples available for training of models leads to further increases with models that have raw infra-red images as inputs (IR, IR+IR and IR+OF models), whereas degradation of performance is seen where only optical flow is the input (TV-L^1 and SPyNet models). This observation is inline with those of Ng et al. [39] where they conclude the decrease is because of insufficient detail in the optical flow images which is exacerbated by cropping or augmentation.

Author Contributions: RS proposed the use of optical flow and using multiple frames to improve the classification. RS carried out the experiments and wrote the draft article. FD and BM oversaw and gave feedback on the drafts, GF gave feedback on the drafts and further suggestions for improvement.

Funding: This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6049 and through MaREI, the SFI Research Centre for Energy, Climate, and Marine (Grant number 12/RC/2302). We also wish to acknowledge the European Research Council (ERC-2019-AdG 833125-HIGHWAVE) and Insight, the SFI Research Centre for Data Analytics (Grant number SFI/12/RC/2289).

Data Availability: The code used to produce these results is available from github.com/ryan597/Precomputation-of-features--classification. The data and extracted features for all models using the data split method in section 3.2 is available at doi.org/10.5281/zenodo.5361958. The original dataset split used in [6] is also freely available at github.com/dbuscombe-usgs/IR_waveclass.

Acknowledgments: The authors would like to thank Jean-Michel Morel for his advice throughout the project.

Conflicts of Interest: The authors declare no conflict of interest.

1. Deike, L.; Popinet, S.; Melville, W. Capillary effects on wave breaking. *Journal of Fluid Mechanics* **2015**, *769*, 541–569. doi: https://doi.org/10.1017/jfm.2015.103.
2. Deike, L.; Lenain, L.; Melville, W. Air entrainment by breaking waves. *Geophysical Research Letters* **2017**, *44*, 3779–3787. doi: https://doi.org/10.1002/2017GL072883.
3. Lubin, P.; Chanson, H. Are breaking waves, bores, surges and jumps the same flow? *Environmental Fluid Mechanics* **2017**, *17*, 47–77. doi: https://doi.org/10.1007/s10652-016-9475-y.
4. Lim, H.; Chang, K.; Huang, Z.; Na, B. Experimental study on plunging breaking waves in deep water. *Journal of Geophysical Research: Oceans* **2015**, *120*, 2007–2049. doi: https://doi.org/10.1002/2014JC010269.
5. Shi, R.; Leng, X.; Chanson, H. On optical flow techniques applied to breaking surges. *Flow Measurements and Instrumentation* **2020**, *72*. doi: https://doi.org/10.1016/j.flowmeasinst.2020.101710.
6. Buscombe, D.; Carini, R. A data driven approach to classifying wave breaking in infrared imagery. *Remote Sensing* **2019**, *11*, 859. doi: https://doi.org/10.3390/rs11070859.
7. Choi, H.; others. Real-time significant wave height estimations from raw ocean images based on 2D and 3D deep neural networks. *Ocean Engineering* **2020**, *201*. doi: https://doi.org/10.1016/j.oceaneng.2020.107129.
8. Horn, B.; Schunck, B. Determining optical flow. *Artificial Intelligence* **1981**, *17*, 185–203. doi: http://dx.doi.org/10.1016/0004-3702(93)90173-9.
9. Sánchez, J.; Meinhardt-Llopis, E.; Facciolo, G. TV-L1 Optical flow estimation. *Image Processing On Line* **2013**, *3*, 137–150. doi: https://doi.org/10.5201/ipol.2013.26.
10. Fukami, K.; Fukagata, K.; Taira, K. Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics* **2019**, *870*, 106–120. doi: https://doi.org/10.1017/jfm.2019.238.
11. Brunton, S.; Noack, B.; Koumoutsakos, P. Machine Learning for Fluid Mechanics. *Annual Review of Fluid Mechanics* **2020**, *52*, 477–508. doi: https://doi.org/10.1146/annurev-fluid-010719-060214.
12. Benetazzo, A. Measurements of short water waves using stereo matched image sequences. *Coastal Engineering* **2006**, *53*, 1013–1032. doi: https://doi.org/10.1016/j.coastaleng.2006.06.012.
13. Fedele, F.; Benetazzo, A.; Gallego, G.; Shih, P.C.; Yezzi, A.; Barbariol, F.; Ardhuin, F. Space-time measurements of oceanic sea states. *Ocean Modelling* **2013**, *70*, 103–115. doi: https://doi.org/10.1016/j.ocemod.2013.01.001.
14. Gallego, G.; Yezzi, A.; Fedele, F.; Benetazzo, A. A Variational Stereo Method for the Three-Dimensional Reconstruction of Ocean Waves. *IEEE Transactions on Geoscience and Remote Sensing* **2011**, *49*, 4445–4457. doi: https://doi.org/10.1109/TGRS.2011.2150230.
15. Stringari, C.; Harris, D.; Power, H. A novel machine learning algorithm for tracking remotely sensed waves in the surf zone. *Coastal Engineering* **2019**, *147*, 149–158. doi: https://doi.org/10.1016/j.coastaleng.2019.02.002.
16. Stringari, C.; Guimarães, P.; Filipot, J.; Leckler, F.; Duarte, R. Deep neural networks for active wave breaking classification. *Scientific Reports* **2021**, *11*, 3604. doi: https://doi.org/10.1038/s41598-021-83188-y.

17. Stringari, C.; Filipot, J.; Leckler, F.; Duarte, R. Remote Sensing Observations of Dominant Breaking Waves in Intermediate to Deep Water from a Lighthouse During Storm Conditions. *Preprints* **2021**. doi: https://doi.org/10.20944/preprints202103.0538.v1.
18. Choi, H.; Park, M.; Son, G.; Jeong, J.; Park, J.; Mo, K.; Kang, P. Real-time significant wave height estimation from raw ocean images based on 2D and 3D deep neural networks. *Ocean Engineering* **2020**, *201*, 107129. doi: https://doi.org/10.1016/j.oceaneng.2020.107129.
19. Buscombe, D.; Carini, R.; Harrison, S.; Chickadel, C.; Warrick, J. Optical wave gauging using deep neural networks. *Coastal Engineering* **2020**, *155*, 103593. doi: https://doi.org/10.1016/j.coastaleng.2019.103593.
20. Jessup, A.; Zappa, C.; Yeh, H. Defining and quantifying microscale wave breaking with infrared imagery. *Journal of Geophysical Research* **1997**, *102*, 23145– 23153. doi: https://doi.org/10.1029/97JC01449.
21. Jessup, A.; Zappa, C.; Loewen, M.; Hesany, V. Infrared remote sensing of breaking waves. *Nature* **1997**, *385*, 52–55. doi: https://doi.org/10.1038/385052a0.
22. Zhang, G.; Chanson, H. Application of local optical flow methods to high-velocity free-surface flows: Validation and application to stepped chutes. *Experimental Thermal and Fluid Science* **2018**, *90*, 186–199. doi: https://doi.org/10.1016/j.expthermflusci.2017.09.010.
23. Quénot, G.; Pakleza, J.; Kowalewski, T. Particle image velocimetry with optical flow. *Experiments in Fluids* **1998**, *25*, 177–189. doi: https://doi.org/10.1007/s003480050222.
24. Corpetti, T.; Heitz, D.; Arroyo, G.; Mémin, E. Fluid experimental flow estimation based on an optical-flow scheme. *Experiments in Fluids* **2006**, *40*, 80–97. doi: https://doi.org/10.1007/s00348-005-0048-y.
25. Heitz, D.; Mémin, E.; Schnörr, C. Variational fluid flow measurements from image sequences: synopsis and perspectives. *Experiments in Fluids, Springer* **2010**, *48*, 369–393. doi: https://doi.org/10.1007/s00348-009-0778-3.
26. Liu, T.; Shen, L. Fluid flow and optical flow. *Journal of Fluid Mechanics* **2008**, *614*, 253–291. doi: https://doi.org/10.1017/S0022112008003273.
27. Wang, B.; Cai, Z.; Shen, L.; Liu, T. An analysis of physics-based optical flow. *Journal of Computational and Applied Mathematics* **2015**, *276*, 62–80. doi: https://doi.org/10.1016/j.cam.2014.08.020.
28. Ng, J.H.; Hausknecht, M.; Vijayanarasimhan, S.; Vinyals, O.; Monga, R.; Toderici, G. Beyond Short Snippets: Deep Networks for Video Classification. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **2015**. https://arxiv.org/abs/1503.08909v2.
29. Sonka, M.; Hlavac, V.; Boyle, R. *Image processing, analysis and machine vision*; PWS publishing, 1999.
30. Zach, C.; Pock, T.; Bischof, H. A duality based approach for realtime TV-L1 optical flow. *Pattern Recognition DAGM 2007. Lecture Notes in Computer Science* **2007**, *4713*, 214–223. doi: https://doi.org/10.1007/978-3-540-74936-3_22.
31. Wedel, A.; Pock, T.; Zach, C.; Bischof, H.; Cremers, D. An improved algorithm for TV-L1 optical flow. *Statistical and Geometrical Approaches to Visual Motion Analysis. Lecture Notes in Computer Science* **2009**, *5604*, 23–45. doi: https://doi.org/10.1007/978-3-642-03061-1_2.
32. Monzón, N.; Salgado, A.; Sánchez, J. Robust Discontinuity Preserving Optical Flow Methods. *Image Processing On Line* **2016**, *6*, 165–182. doi: https://doi.org/10.5201/ipol.2016.172.
33. Ranjan, A.; Black, M. Optical Flow Estimation using a Spatial Pyramid Network. *IEEE Conference on Computer Vision and Pattern Recognition* **2017**.
34. Niklaus, S. A Reimplementation of SPyNet Using PyTorch. https://github.com/sniklaus/pytorch-spynet, 2018.
35. Youden, W. Index for rating diagnostic tests. *Cancer* **1950**, *3*. doi: https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
36. Brier, G. Verification of forecasts expressed in termps of probability. *Monthly Weather Review* **1950**, *78*. doi: https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.
37. Howard, A.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**. https://arxiv.org/abs/1704.04861.
38. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2017**, p. 1251–1258.
39. Ng, Y.H.; others. Beyond Short Snippets: Deep Networks for Video Classification. *Computer Vision Foundation* **2015**, pp. 4694–4702.