HAL
open science

Detection of point scatterers using diffraction imaging and deep learning

Valentin Tschannen, Norman Ettrich, Matthias Delescluse, Janis Keuper

To cite this version:

Valentin Tschannen, Norman Ettrich, Matthias Delescluse, Janis Keuper. Detection of point scatterers using diffraction imaging and deep learning. Geophysical Prospecting, 2019, 10.1111/1365-2478.12889 . hal-02354803

HAL Id: hal-02354803

https://hal.science/hal-02354803v1

Submitted on 7 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Geophysical Prospecting EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Geophysical Prospecting, 2019 doi: 10.1111/1365-2478.12889

Detection of point scatterers using diffraction imaging and deep learning

Valentin Tschannen[1,2]*, Norman Ettrich[1], Matthias Delescluse[2] and Janis Keuper[1]

[1]*Fraunhofer Center for Machine Learning and Industrial Mathematics (ITWM), Competence Center for High Performance Computing, 67 663, Kaiserslautern, Germany, and* [2]*École Normale Supérieure, UMR 8538, PSL Research University, 75231, Paris, France*

Received March 2019, revision accepted October 2019

ABSTRACT

Diffracted waves carry high-resolution information that can help interpreting fine structural details at a scale smaller than the seismic wavelength. However, the diffraction energy tends to be weak compared to the reflected energy and is also sensitive to inaccuracies in the migration velocity, making the identification of its signal challenging. In this work, we present an innovative workflow to automatically detect scattering points in the migration dip angle domain using deep learning. By taking advantage of the different kinematic properties of reflected and diffracted waves, we separate the two types of signals by migrating the seismic amplitudes to dip angle gathers using prestack depth imaging in the local angle domain. Convolutional neural networks are a class of deep learning algorithms able to learn to extract spatial information about the data in order to identify its characteristics. They have now become the method of choice to solve supervised pattern recognition problems. In this work, we use wave equation modelling to create a large and diversified dataset of synthetic examples to train a network into identifying the probable position of scattering objects in the subsurface. After giving an intuitive introduction to diffraction imaging and deep learning and discussing some of the pitfalls of the methods, we evaluate the trained network on field data and demonstrate the validity and good generalization performance of our algorithm. We successfully identify with a high-accuracy and high-resolution diffraction points, including those which have a low signal to noise and reflection ratio. We also show how our method allows us to quickly scan through high dimensional data consisting of several versions of a dataset migrated with a range of velocities to overcome the strong effect of incorrect migration velocity on the diffraction signal.

Key words: Seismics, Imaging, Modelling, Signal processing.

INTRODUCTION

Most of the information obtained in exploration seismology comes from specular energy associated with reflections of acoustic and elastic waves at boundaries between geological layers. The resolution of the images is limited by the bandwidth of the seismic wavelet and, in particular, objects whose size is small compared to the dominant wavelength cannot be well resolved. Many small-scale structural details such as reflector discontinuities at fault planes, karst, pinch-outs, channel edges, boulders or sand injectites may fall below the resolution power of reflection images and undermine the quality of the interpretation. Yet these objects will provoke the scattering of waves in all directions, a phenomenon called

*Email: valentin.tschannen@itwm.fraunhofer.de

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

diffraction. Due to their truncated Fresnel zone, diffracted waves contain a higher resolution signal and are valuable for an enhanced interpretation and inversion (Khaidukov, Landa and Moser 2004; Moser and Howard 2008; Landa, Fomel and Reshef 2008; Huang, Zhang and Schuster 2015). However, diffracted signal is not often available to the interpreter because its amplitude can be far weaker than reflected signal and is often lost during processing (Khaidukov *et al.* 2004).

It is possible to separate reflected signal from specular signal by taking advantage of the different kinematic properties of the waves interactions (Landa, Shtivelman and Gelchinsky 1987; Kanasewich and Phadke 1988). When encountering a locally planar surface, the energy will reflect in a focused direction depending on the angle of incidence. On the other hand, a diffractor will scatter the energy in all directions. In particular, this distinction will strongly affect the appearance of both signals in the migration-dip angle domain, as this angle is associated with the illumination direction. Using ray-based prestack depth migration in the local angle domain, we can form common image gathers (CIGs) in dependence on the migration dip angle. When computing the diffraction response at a migration point located exactly on a scattering object, given that the correct velocity model is used, the migration operator will align with the diffraction hyperbola and give rise to a flat response in the dip angle CIG. In contrast, when migrating a point located on a reflector, the operator will respond only at a dip corresponding to the structural inclination angle (Audebert *et al.* 2005; Landa *et al.* 2008; Reshef and Landa 2009; Klokov, Baina and Landa 2010; Klokov and Fomel 2012).

Several authors have suggested algorithms to detect scattering points from seismic data. Klokov and Fomel (2012) use a hybrid Radon transform to detect diffracted waves from time domain dip angle CIGs. Arora and Tsvankin (2017) show that discrimination in the dip angle domain is also possible in a transversely isotropic media. Shustak and Landa (2017) and Dafni and Symes (2017) form dip angle CIGs using reverse time migration, making their method applicable in complex geological areas exhibiting strong local variations in the velocities.

In order to detect the diffraction points, those approaches usually rely on filtering out the dips around the estimated specular dip before stacking followed by a visual inspection of the seismic image. This may be a time consuming and also challenging task due to the weak energy of the diffraction signal. For the same reasons, it is challenging to design a detection filter that will be reliable especially in areas of low signal to noise ratio. However, it has now become common practice to resort to supervised machine learning, a branch of artificial intelligence, to solve pattern recognition problems. Rather than requiring a hand crafting of the detection function, this class of algorithms is based on the use of a number of free parameters that will learn from examples. In particular, most of the attention is currently received by the field of deep learning carried by the so-called deep neural networks algorithms that have become the state of the art in solving a variety of recognition tasks (LeCun, Bengio and Hinton 2015). Several authors already applied deep learning to automatically detect structural features in seismic data. Waldeland and Solberg (2017) and Guitton (2018) employed a convolutional neural network (CNN) to, respectively, detect salt bodies and faults from a stack. Pham, Fomel and Dunlap (2018) used a pixel-wise CNN to locate channels networks. Regarding diffraction, de Figueiredo *et al.* (2013) suggested to use a machine learning based k-nearest neighbours classifier to detect diffracted signal from common offset gathers and applied their method on ground penetration radar data. Serfaty *et al.* (2017) suggested to distinguish diffracted events from other signals by working with compressed gathers in the local angle domain. After manually labelling a small number of seismic patches according to the dominant object they contain (reflector, fault, point diffractor, migration noise, random noise), they used a network pre-trained on natural images and retrained the last classification layer. The authors show that their method was successful in identifying diffracted waves in their dataset.

While deep neural networks are the best classifiers currently known, resorting to them comes with several difficulties. Probably the most challenging part is to create a training dataset. Not only do we need the seismic amplitudes but also the labels indicating the correct interpretation of those data. In practical applications on natural images, state-of-the-art results usually require to train the networks on millions of examples (LeCun *et al.* 2015). Creating such a dataset, especially in geoscience where expertise is required for the interpretation and where uncertainty and subjectivity will lead to non-unique answers, present a serious challenge. Additionally, the networks are made of many, sometimes millions, of parameters and a number of key hyper-parameters with complex non-linear dependencies. Finding the correct settings is both computationally and manually time consuming. To judge the quality of the obtained solution, authors usually evaluate performances with a blind test on a subset of the dataset that was not used during training. However, in the case where this data subset is statistically similar to the training data, a good test accuracy will not necessarily mean good generalization

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

performances (LeCun *et al.* 2015). In that case, the network might fail to produce an acceptable answer on a new dataset with a different geology or different processing.

In this work, we present an automated workflow that does not require manual picking of the diffracted waves. We use modelling to design a large synthetic dataset with low interpretation uncertainties and train a CNN from full dip angle gathers to perform high-resolution detection of scattering objects. In particular, we are aiming to identify diffracted signal emitted by point scatterers, that is objects which have a compact shape in all three spatial dimensions. Diffraction may also occur at edges, yielding a different pattern (Trorey 1970), that we do not treated in this work. After evaluation on synthetic data, we show that our network is able to identify diffracted events on a field-recorded dataset, yielding encouraging results as for its generalization capabilities. The first part of our paper gives the reader elements for an intuitive understanding of the diffraction imaging process and the training of a deep neural network. We then give practical details about the synthetic data creation and the training process. Finally we perform a case study on real data and discuss the current limitations of the method as well as further research opportunities.

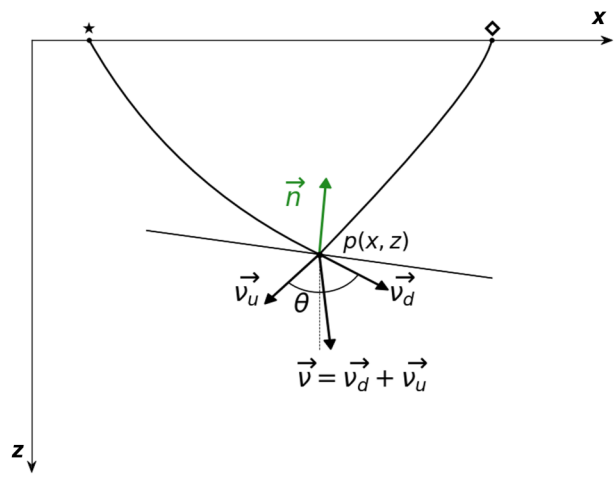

Figure 1 Two-dimensional representation of the local angle domain imaging geometry. A ray pair obtained by shooting from the migration point $p(x, z)$ and reaching a source/receiver pair is drawn. Vectors $\boldsymbol{v}_d$ and $\boldsymbol{v}_u$ are the tangents to the slowness vectors of the down- and up-going rays at p. The dip vector $\boldsymbol{v}$ is defined as the sum of those vectors, and migration dip ν is the angle between $\boldsymbol{v}$ and the vertical. The opening angle θ is the angle between $\boldsymbol{v}_d$ and $\boldsymbol{v}_u$. $\boldsymbol{n}$ is the normal to the locally planar geological reflector.

METHODS

Diffraction imaging

Prestack depth migration aims at mapping seismic events from the acquisition domain to their true position in the subsurface. This method requires the knowledge of an accurate spatially varying velocity field as well as a technique to simulate the propagation and back-propagation of waves. Under the assumption that the dominant wavelength of the seismic wavelet is much smaller than the scale of heterogeneities of the velocity model, we can replace the direct integration of the wave equation by a lighter ray tracing algorithm to compute the propagation. The algorithm we used in our study is working in the local angle domain, treating every node of the imaging grid as a point scatterer (Audebert *et al.* 2005; Moser and Howard 2008; Merten and Ettrich 2015). In the following, we explain the procedure in more detail without accounting for implementation performance specificities. In a two-dimensional (2D) acquisition, we can sort the data as common source gathers where seismic events are located by the source and receiver positions along the sailing line x and the two-way travel time t of the waves. We define our image space with a discrete grid parameterized by x and the depth z and build the migrated image iteratively at every node $p(x, z)$ of the subsurface model. We treat p as a scattering point and start by shooting a ray fan through the velocity model to form ray pairs that, respectively, reach existing source–receiver couples. Figure 1 shows the parameterization of the ray pairs in terms of the opening angle θ and the migration dip angle ν. Then, we compute the travel times along every ray pairs contributing to p (as well as an amplitude correction factor that accounts in particular for the geometrical spreading of the energy along the expanding wavefront) and fetch the corresponding samples in the shot gather. At this stage, for a given velocity field, the migrated data will be depending on four parameters: its position in space given by x and z and the local imaging angles ν and θ.

As the dip angle is directly associated with the illumination direction, reflected and diffracted events will exhibit distinct responses in the dip angle domain. By summing the data along opening angles, dip angle common image gathers (CIGs) are formed by constructive interferences (Landa *et al.* 2008; Klokov *et al.* 2010). Figure 2 illustrates the dip angle response of both a horizontal reflector and a scattering point. For the horizontal reflector, the zero-offset recorded wavefield will show a seismic event at a constant time for every position on the sailing line. Points p_0, p_1 and p_2 represent three migration nodes on a vertical line located at $x = x_k$. From their corresponding migration operators, we observe that nodes

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers., *Geophysical Prospecting*, 1–15

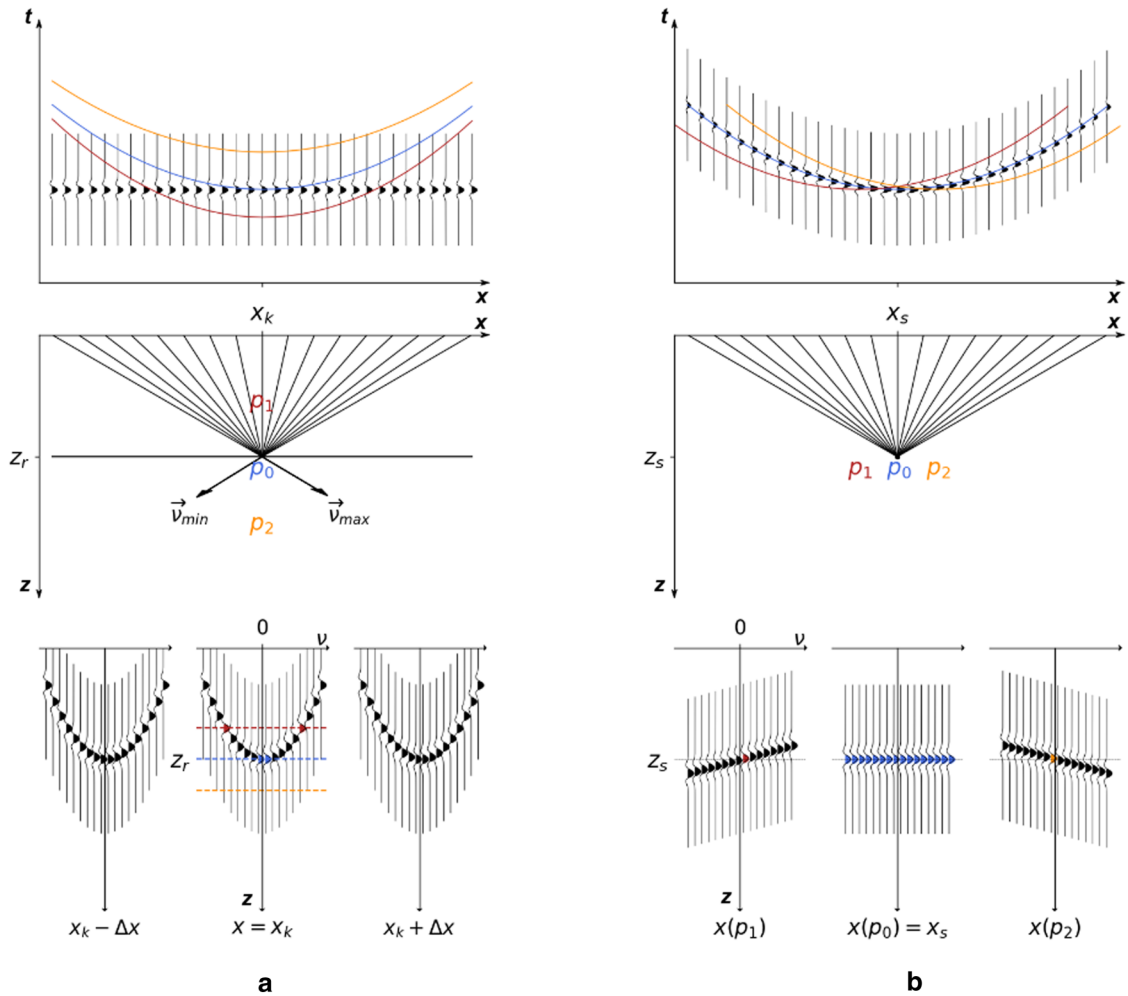

Figure 2 Two-dimensional illustration of the dip angle response, drawn for the zero-offset case, of (a) a horizontal planar reflector at depth $z = z_r$ separating two constant velocity half-spaces (b) a point scatterer in a constant velocity space located in (x_s, z_s). The central row represents the subsurface model. p_0, p_1 and p_2 are migration points. Ray pairs are plotted for p_0 as well as the corresponding minimum and maximum dip-vectors (see Fig. 1). The upper row shows the recorded wave field. Colour-coded diffraction curves of the three migration points are displayed on the seismograms. The bottom panel represents the migrated wavefield sorted as dip angle gathers. Coloured wiggles correspond to the amplitudes picked by the migration operators.

beneath p_0 will not contribute to the imaging of the reflector. The diffraction operator of the node p_0, which is located exactly on the reflector, will encounter the signal only at its apex corresponding to the zero dip angle. Migration operators of the nodes above p_0 will cross the wiggles at positions progressively further away from their apex and contribute to the imaging at progressively larger absolute dips, leading to a parabola-like shape of the signal in the gather. One should note that in the case of a tilted planar reflector, the apex of this pseudo-parabola will be at the dip equal to the plane's normal direction (vector $\boldsymbol{n}$ in Fig. 1). For the point scatterer of Figure 2, things are different as the waves will no longer reflect in a focused direction but will be scattered in all directions. The migration operator computed at p_0, positioned on the diffraction point, will match exactly the diffraction hyperbola of the recorded wavefield. Seismic events will be fetched at every dip angle, creating a flat horizontal response in the gather. When migrating a vertical line on the left of the scattering object, migration surfaces will cross the diffraction hyperbola at dips progressively increasing as the depth of the nodes decreases (and inversely for a line on the right-hand side of the diffractor), creating a flat tilted signature in the gathers. The slope of the response will increase as the lateral position of the line gets away from the diffractor. Similar observations hold

for non-zero offset data. Moreover, we see with this example that even for a short offset range large range dip gathers can be created.

The focusing quality of the method relies on having been able to derive an accurate velocity model. Figure 3 illustrates the sensitivity of the diffracted signal to errors in the velocity field. It is possible to use this strong sensitivity to perform high-resolution velocity analysis (Sava, Biondi and tgen 2005; Fomel, Landa and Taner 2007).

The relatively weak amplitudes of the diffracted waves combined with their strong sensitivity to accurate velocity models and pre-processing steps make their interpretation delicate. The interpreter will need to look through a large amount of prestack data, and potentially through several migrated versions of the same dataset obtained with different velocity models. Additionally, certain geological areas might be very rich in scattering objects, making their identification a time consuming and tedious task. For these reasons, it is preferable to employ an automated method, robust to low signal to noise ratio, and yielding a high-resolution detection. In the next section, we introduce the use of deep learning to solve such problem.

Semantic segmentation

Semantic segmentation deals with the problem of classifying every singular pixel of the data among a set of classes. This approach is a high-resolution extension of data classification, which aims at associating a class label to a group of pixels. Supervised deep learning has now become the method of choice to tackle those problems as it has proven to bring best performances on a broad range of applications (LeCun *et al.* 2015). The main technology behind deep learning is the so-called neural networks. These networks are built as a sequence of layers forming a non-linear, piece-wise differentiable function connecting the input to the output. Every layer is responsible for performing a simple affine transformation on its input and applies an element-wise non-linearity. The power of those networks resides in the way the parameters of the transforms are set. Rather than being manually engineered, the parameters are initially chosen at random and given the freedom to automatically adapt to the data by progressively learning from examples. Stacking several layers is a key to the success of those algorithms since this architectural design allows them to learn a hierarchical representation of the data. The deeper layers will benefit from the work of the previous layers and will be sensitive to progressively more abstract and complex features expressed as a composition of the simpler features learnt by the shallower layers (LeCun *et al.* 2015). Such networks can in theory approximate any function (Cybenko 1989). When the data exhibits a spatial structure and the surrounding information is relevant to understand the local context, a suited choice for the linear transformations is convolutions, and the family of algorithms based on them is called convolutional neural networks (CNNs) (LeCun *et al.* 1998).

In this work, we want to identify and localize diffraction points in the subsurface using seismic data. Learning to identify those elements consists in optimizing a neural network in order to approximate the distribution $\mathcal{D}$ over the domain $S = H \times P$, where H is the space of seismic data and P is the space of probabilities indicating the likelihood for the presence of diffraction points. Our training dataset consists of a collection of patches $d_1, d_2, \ldots, d_N \in S$ drawn from $\mathcal{D}$. In the 2D case, for a given $d \in S$, $d = \{\mathbf{h}^\alpha, \mathbf{p}\}$ is a tuple formed of a prestack seismic amplitudes patch $\mathbf{h}^\alpha(x, z, \nu)$ and its corresponding mask of probable locations of the scattering points $\mathbf{p}(x, z)$ (see Fig. 4 for an example). In deep learning, we refer to the different prestack sections as channels and call a single channel section a feature map. As an example, when working on natural images, the input data contains three channels, forming a coloured image as a composition of the red, green and blue feature maps. In our binary problem, either there is a diffraction point in this pixel or there is not, the mask $\mathbf{p}$ is defining at every spatial sample the probability mass function $(p, 1 - p)$, where $0 \leq p \leq 1$ is the probability of having a scattering point.

Figure 4 is a schematic representation of the CNN architecture we use in this work. The network accepts as input prestack seismic data that it will progressively transform and reshape in order to output a patch matching the shape of the mask. It is composed of four types of layers in charge of performing convolutions, down-sampling, up-sampling and softmax scaling. The trainable parameters are the kernels (and biases) of the convolutional layers. We illustrate in Figure 5 how the first convolutional layer is working. A single input data sample $\mathbf{h}^\alpha = (h_1^\alpha, \ldots, h_{n_\alpha}^\alpha)$ is represented as the concatenation in the channel dimension of n_α feature maps. In our work, n_α is the number of migration dips ν. The output of this layer $\mathbf{h}^\beta = (h_1^\beta, \ldots, h_{n_\beta}^\beta)$ is represented as the concatenation in the channel dimension of n_β feature maps. n_β is an architecture hyper-parameter and corresponds to the chosen number of convolution kernels of the first layer. Every single feature map of $\mathbf{h}^\beta$ is obtained by convolving the input data with a different kernel. Equation (1) expresses the exact operation performed:

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers., *Geophysical Prospecting*, 1–15

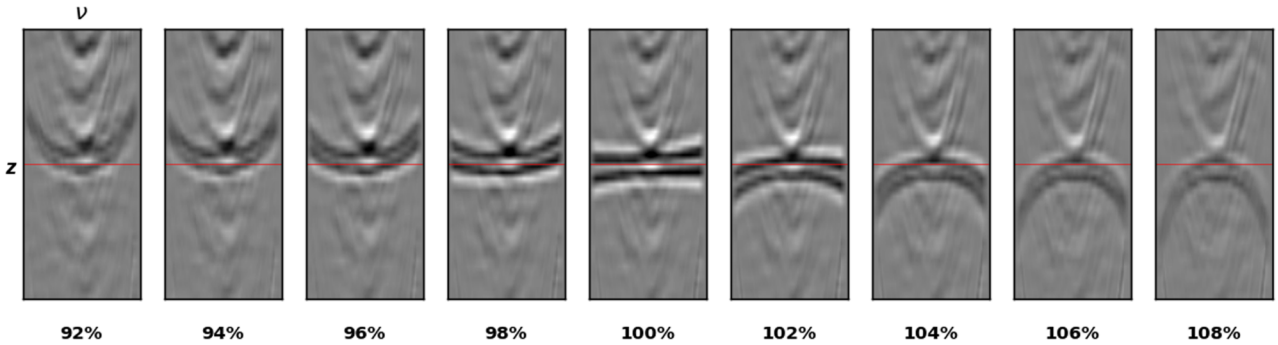

Figure 3 Sensitivity of the diffraction signal to errors in the migration velocity. Every image is a dip angle gather extracted at the same x coordinate located above a synthetic diffraction point. 100% corresponds to the true migration velocity, while remaining percentages correspond to relative perturbations from 2% to 8%.

$$h_i^\beta = \gamma\left(\sum_{k=1}^{n_\alpha} h_k^\alpha * w_k^i + b^i\right), \quad i = 1,..,n_\beta. \tag{1}$$

In the 2D case, the ith weight $\mathbf{w}^i = (w_1^i, \ldots, w_{n_\alpha}^i)$ is a prestack 2D kernel containing the same number of channels as the input layer. A 2D convolution $*$ is performed independently for every channel, and the results are then summed across channels. The ith bias term b^i is added after summation to enable the linear transformation performed by the convolution to be translated from the origin. An element-wise non-linear operator $\gamma(.)$, called the activation function, is applied to break the linearity between the layers in order to increase the approximation capabilities of the network. By repeating equation (1) with n_β weights $(\mathbf{w}^i)_{i=1,\ldots,n_\beta}$ and concatenating the resulting feature maps, we create the new input data for the next layer.

In addition to convolutions, the network also performs spatial down-sampling and up-sampling of the feature maps. The down-sampling is achieved by sliding a small spatial window that selects the largest value and drops the

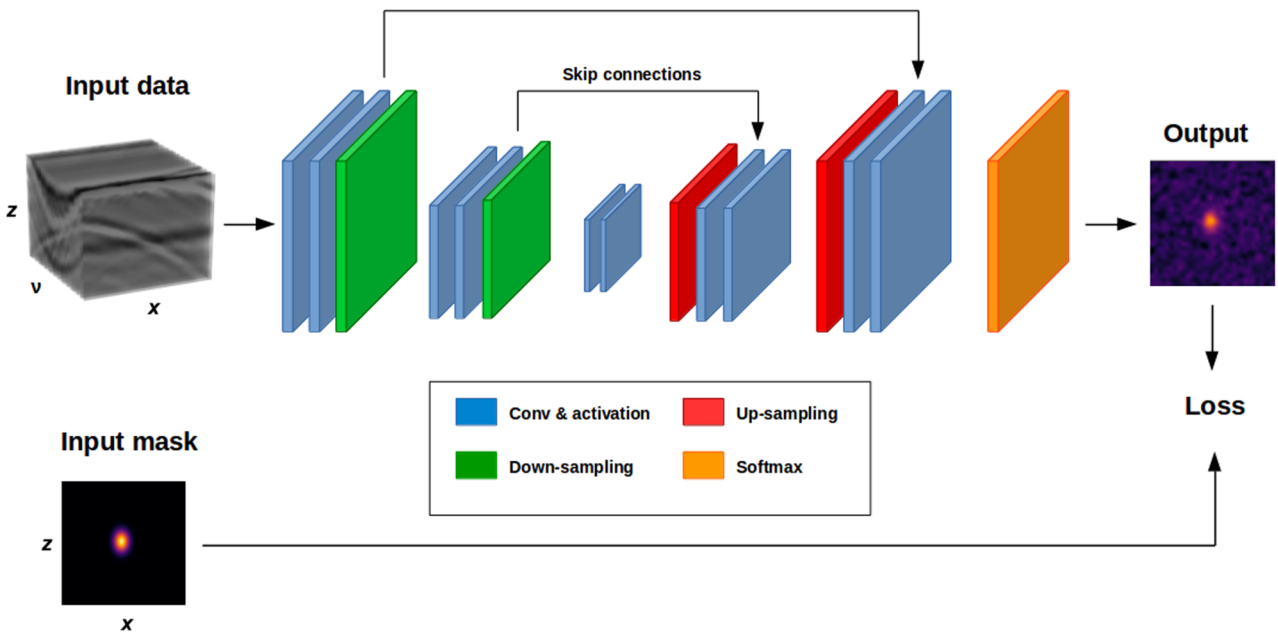

Figure 4 Architecture of our convolutional neural network (designed after Ronneberger, Fischer and Brox 2015). The data flow direction for the forward pass is represented by the black arrows. Input data $\mathbf{h}^\alpha(x, z, v)$ is a 2D prestack data patch and input mask $\mathbf{p}(x, z)$ a 2D patch matching the spatial dimension of the data. Boxes represent multi-channel feature maps colour coded by layer type. Output data $\mathbf{q}(x, z)$ has the same dimension as the input mask.

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

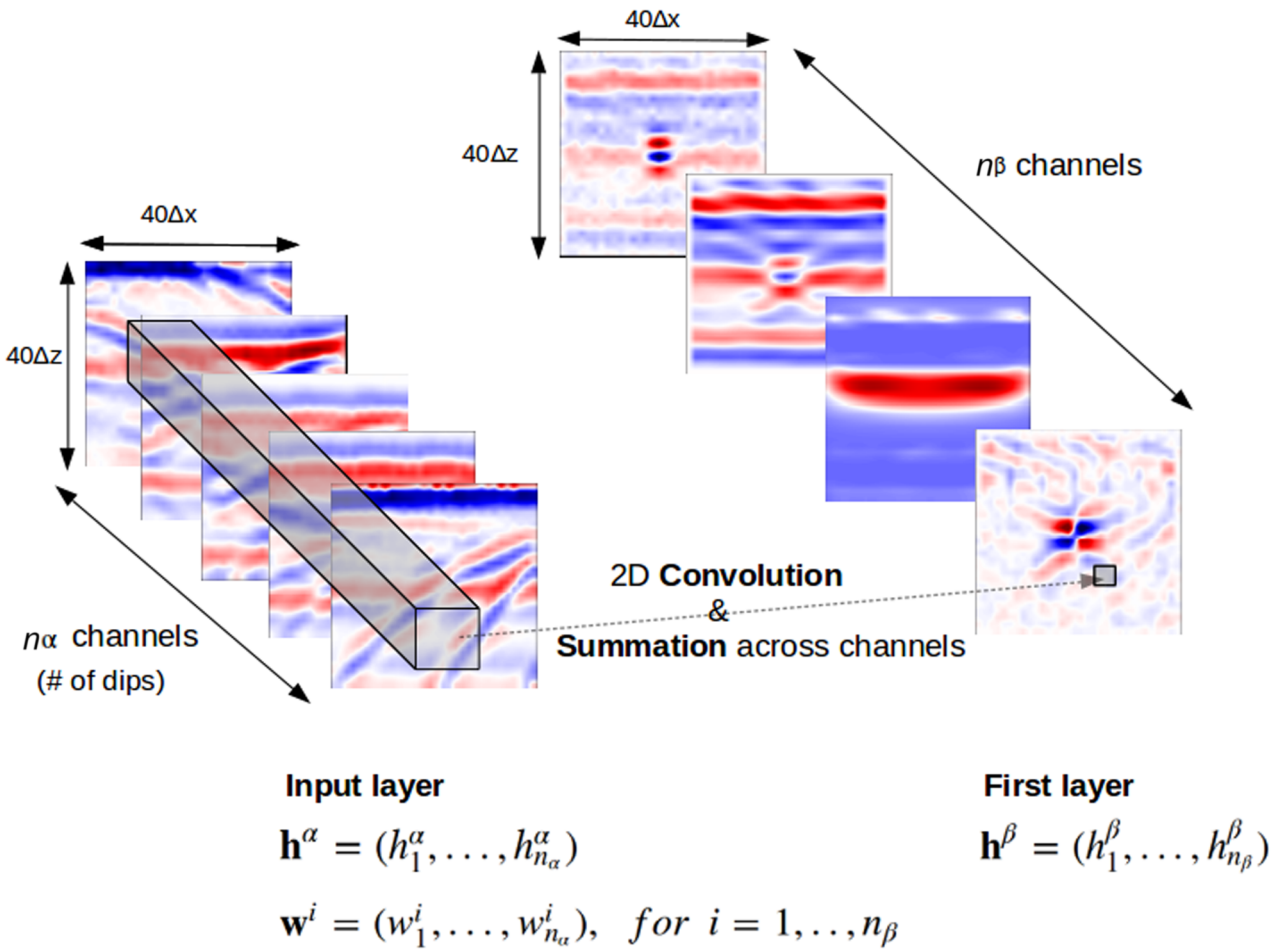

Figure 5 Principle of the first convolutional layer (in the 2D case, without bias adding and activation). The input data is a prestack 2D seismic patch $\mathbf{h}^{\alpha}(x, z, \nu)$ of shape $40\Delta x \times 40\Delta z$ with n_α channels corresponding to the number of dips. The kernels $\mathbf{w}^i$ have the same rank as the input data, and the 2D convolution is performed in the space (x, z) for every channel. An example of prestack kernel is overlay on the input seismic. After the convolution, a 2D output feature map is obtained by summing across channels. Every feature map h_i^{β} is obtained after convolving with a different kernel. The output data $\mathbf{h}^{\beta}$ is formed by the concatenation of the n_β feature maps.

remaining ones. While increasing the non-linearity of the network and forcing translation invariance, this operation also has the effect of expanding the receptive field of the convolutional kernels. By using a small constant spatial shape (e.g. 3×3) through every layer, down-sampling enables the kernels to progressively access a larger area of the data. This characteristic is important in order to learn a multi-scale representation, developing the abstraction power of the network. The up-sampling is the reverse operation and is used to progressively bring back the feature maps to their original spatial shape, which is a requirement since our network should perform a pixel-wise classification. A common technique for up-sampling is to learn the operation using strided transpose convolutions (Long, Shelhamer and Darrell 2014). A final architectural specificity is the use of skip connections (Fig. 4; Ronneberger, Fischer and Brox 2015) to reuse data from shallow layers in the deepest ones. Shallow feature maps are concatenated along the channels with their deeper counterpart of identical spatial size. This will allow the network to make use of both feature maps coming from early layers that contain information close to the original data and feature maps from later layers that contained highly transformed information. In order to map the output of the network to a pseudo-probability distribution expressing the classes diffraction and non-diffraction points, we design the output to be composed of two feature maps and scale every prestack pixels using a softmax layer (e.g. LeCun *et al.* 1998; Fig. 4). The output $\mathbf{q}(x, z)$ is defining at every spatial samples a mass function

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers., *Geophysical Prospecting*, 1–15

$(q, 1-q)$, where $0 \leq q \leq 1$ indicates the confidence of the network in having found a diffraction point.

We called $\mathcal{D}$ the unknown *true* distribution that expresses the probability of having diffraction points in the seismic data, and let $\hat{\mathcal{D}}$ be the distribution computed by our network. Training a neural network consists in optimizing the values of its parameters $w \in \mathbb{R}^m$, where m is the number of free dimensions, in order to increase its prediction performances by bringing $\hat{\mathcal{D}}$ close to $\mathcal{D}$. For a given data point $d \in S$, we evaluate the quality of the prediction by the error measure $l(w; d)$ and define the training procedure as the minimization of the loss function over the finite dataset S:

$$\mathcal{L}(w; S) = \frac{1}{N} \sum_{i=1}^{N} l(w; d). \tag{2}$$

Given a training sample $d = \{\mathbf{h}^{\alpha}, \mathbf{p}\}$, we are concerned with minimizing the error of the network prediction $\mathbf{q}$. A standard pseudo-distance measure between two probability distributions is the cross-entropy (e.g. LeCun *et al.* 1998). It will measure how close is the computed distribution in representing the *true* distribution. In its binary form, the cross-entropy between $\mathbf{p}$ and $\mathbf{q}$ can be expressed as

$$l(w; d) = -p \log(q) - (1-p) \log(1-q). \tag{3}$$

Equation (3) shows that cross-entropy is differentiable and convex with respect to $\mathbf{q}$ (but not necessary convex with respect to w), and its minimum is reached at $q = p$. So far, the only computationally tractable way to minimize the loss function of equation (2) is to use a steepest descent algorithm (LeCun *et al.* 2015). Given a position in the optimization landscape for a parameterization state w_t, the method consists of finding the local downhill direction expressed by the negative gradient of the function computed at that point. A move towards a new point of the landscape is done by updating the parameters of the function in this direction, $w_{t+1} := w_t - \eta \nabla \mathcal{L}(w_t, S)$, where the gradient ∇ is the first-order vector derivative and η, called the learning rate, is the hyper-parameter defining the step size of the descent. Since the loss function directly depends only on the last layer of the network, we need to use the derivative chain rule in order to back-propagate the gradient to earlier layers (Werbos 1974). The procedure is repeated iteratively until convergence to a local minimum is obtained. In practice, it is not feasible to compute the gradients for every points of our training set at once, and we rather use a small random subset of the data, called batch, at every iteration. When every example has been seen once by the network, we say it was trained for one epoch. In addition, it is common to keep a moving average of past gradients and use it to influence the latest decent direction for better performances in the case of ill-conditioned loss landscapes (Rumelhart *et al.* 1988). Put together, this minimization procedure is called momentum stochastic gradient descent.

Coming up with the architecture and set of hyper-parameters that perform well on a given dataset can be a tedious task. Most of the field of deep learning is based on empirical findings, and the time needed to design a network is usually spent on hand tuning a number of parameters in order to increase the performances on the testing data. Moreover, the very high dimensionality of the optimization space combined with its non-convexity might provoke the convergence towards a *bad* local minimum. When this happens, one can achieve very good performances on a certain dataset but the network generalization capability will be poor and therefore lead to incorrect results when evaluated on new data with a non-trivially overlapping statistical distribution. In practice, it seems that to overcome these limitations, one needs to train the network with many, sometimes millions, labelled examples (LeCun *et al.* 2015). In the next section, we expose our strategy to create a training dataset using wave equation modelling.

RESULTS

Training on synthetics

Probably the most challenging part in designing a deep learning based application is not building the algorithm but rather preparing the data that will be used for training and evaluation, and most engineering-level applications of convolutional neural networks (CNNs) require to prepare a very large number of examples to produce robust and generalizable results (LeCun *et al.* 2015). While semantic segmentation offers a high-resolution interpretation of the data, it comes at the cost of having to prepare label masks. Such labels are difficult to get by since we need to annotate every pixel of the training data, and applications to seismic images usually require expertise in order to provide an acceptable interpretation. At this end, rather than manually labelling real data, we resort to synthetic modelling to create training and testing sets. Additionally, because of the inerrant uncertainties on geophysical data, manual labelling is prone to errors and subjectivity, while modelling allows us to use physics to control the procedure. As our approach is fully automated, we can cover for a wide range of velocity contrasts and source wavelets, in order to incorporate as much diversity in our training examples as

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

possible. In the next section, we will evaluate our network performance on high-resolution field-recorded data and we have chosen the modelling parameters accordingly.

We simulated fifty two-dimensional (2D) marine acquisitions with 250 Hz Ricker wavelets using a finite difference integration of the acoustic wave equation on a $\Delta x \times \Delta z = 0.5\,\text{m} \times 0.5\,\text{m}$ grid. Punctual high acoustic impedance perturbations are added to the layered velocity and density models to simulate the diffraction points. Our labels consist of binary masks indicating the position of the diffractors on the grid by a 1 and 0 elsewhere. In order to allow for uncertainties in the exact position of the scattering points, we convolved the masks with a normalized anisotropic 2D Gaussian (see label patch in Fig. 4). The central point of the Gaussian indicates the most likely position, and surrounding pixels show a progressive decay of the likelihood. After migrating the seismic to gathers with dips ranging from -40° to 40°, we created our training dataset by randomly extracting 200,000 prestack patches of shape $40\Delta x \times 40\Delta z$ with the corresponding masks (see Fig. 4 for an example). To augment the diversity of the training data, we also post-processed them with random band-pass frequency filtering and phase rotation. The architecture of our network is presented in Fig. 4. Every convolutional layer contains twenty-four 3×3 kernels. We trained with a momentum stochastic gradient descent optimizer for 30 epochs with a batch size of 48, using an initial learning rate of 10^{-3}. To regularize the training and try to avoid over-fitting, we perturbed the input patches with additive white noise and applied dropout (see, e.g., Ronneberger *et al.* 2015) and a decay factor of 50×37 every 10 epochs to the learning rate.

To control the quality of the training, we additionally created 10 synthetic datasets with a similar method. Figure 6 shows an example of the application of the trained network on this data. To count and localize the diffraction points found by the network, we run a filter on the predicted attribute to find every local maxima. Our parameters are set such that a local maximum should be detected only above a 0.5 confidence and two local maxima should be separated by at least 2 m. We compared the position of those maxima with those of the synthetic perturbations we added to the model and obtained a rate of a 100% true positives and no false negative. We also ran the network on a dataset imaged with a range of velocities to analyse the sensitivity of the prediction to errors in the migration model. Figure 7 illustrates that the algorithm is resilient to small errors in the velocities and is reaching its highest confidence for velocities close to the true one.

While the evaluation on synthetic data shows good performances, one should be careful before extrapolating and claiming that comparable performances will be achieved on any dataset. It is indeed well known that neural networks can easily over-fit the training data without learning to extract meaningful information. Then, if our blind test data are statistically similar to the train data, one is to expect our evaluation metric to yield good results. However, this is not a guaranty that we have solved our problem by creating a robust network that can generalize well. For instance, in this section, we are using synthetic validation data created with a similar approach than the training ones, which might not be enough for a thorough evaluation. A second case is when we use real examples labelled by an interpreter to train the network and evaluate the performances on the same data few hundreds of meters away. It might not be a guaranty that the evaluation metric will still be good if measured on a new dataset with different geology and processing. Another concern deals directly with the evaluation performance measure and training loss we use. Since the interpretation of our data is uncertain, it is unclear what the truth is and trying to match exactly, uncertain, and sometimes wrong labels might be a problem. In our case, it is difficult to know where the scatterers exactly stand in the subsurface and what is their exact spatial extent. In the next section, we judge the performances of our trained network on field-recorded data.

Field data evaluation

The data we use in this study is a 3.5-km line acquired in shallow waters with a high-resolution, shallow penetration source. It serves in a preliminary study to plan the construction of an offshore wind farm. Since the area is a former moraine, it is expected to contain small-scale debris, brought by a glacier, that need to be avoided while drilling.

Figure 8 shows the result given by the network trained on synthetic data. Since we do not know the true number and the location of the scattering objects, we cannot easily give a quantitative measure of the performance of the network. To assess the results, we investigated the data manually. Figures. 9 and 10 show examples of objects found by the network that we believe to be indeed boulders. In Fig. 11, we can see examples of misclassification. A shallow diffraction point with a noisy prestack response was not at all recognized by the machine, while a cross-shaped signal was, we believe, misclassified as a diffraction. Overall, we are satisfied with the rate of true positives. Most areas highlighted by the attribute seem to correspond to actual diffracted events. Estimating the number of false negatives is more difficult, but the dense coverage observed in Fig. 8

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

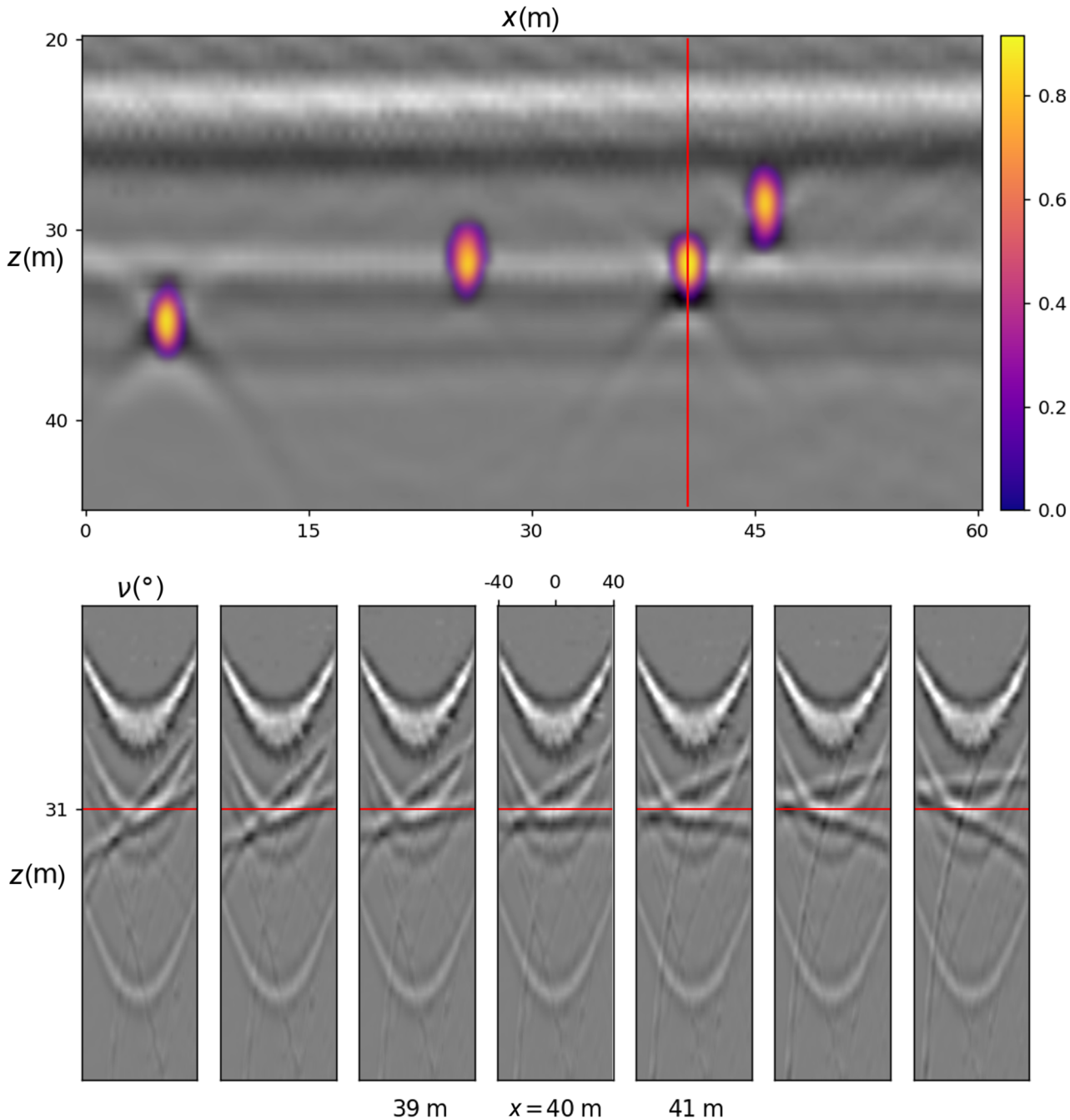

Figure 6 Prediction of the network on synthetic data that was not used for training. The machine confidence in having found a diffraction point is overlaid on the seismic stack. Likelihood inferior to 0.1 is set to be transparent. A vertical red line at $x = 40$ m passes through a diffractor and indicates the location of the central dip-gather displayed at the bottom. Remaining gathers are extracted every meter on the left and right sides. An horizontal line at $z = 31$ m highlights the signature of the scatterer.

gives us confidence that it found a majority of the scattering points.

We tried to incorporate as much diversity as possible in the synthetic data to cover a wide range of possible geologies, but they remain nevertheless a simplification of the reality. Using the confidence attribute, we selected the most probable diffracting objects in the field data and used them to fine-train the network. The results after such training did

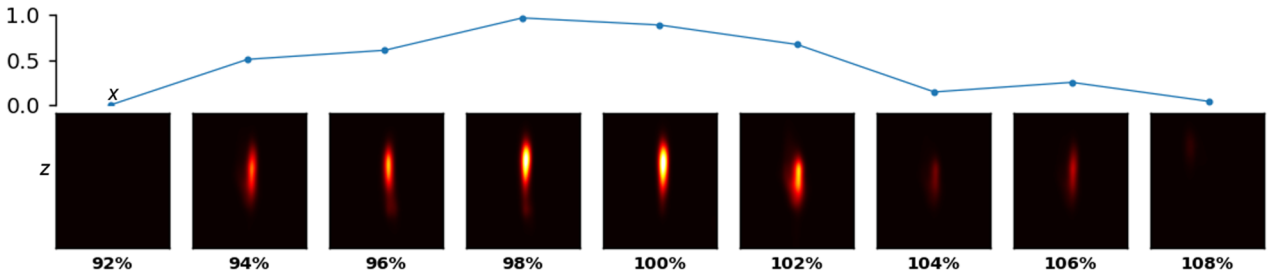

Figure 7 Prediction of the network in a 25 m × 25 m square around a synthetic point scatterer for data migrated with different velocity perturbations as presented in Figure 3. The maximum confidence scores are plotted above their corresponding patch.

not dramatically improve, but we believe that progressively extending our training dataset with real examples found by the network in different datasets will prove useful in further work to increase performances.

DISCUSSION

Mapping seismic amplitudes to the dip angle domain is a useful approach to help separate between diffracted and specular wavefields. Supervised deep learning has become the method of choice to automatically identify patterns in high dimensional data, and in particular convolutional neural networks (CNNs) are a suitable choice for seismic images as they learn, multi-scale, spatial relationships to support their decision. We proposed an automated workflow to detect point scatterers from prestack seismic data using deep learning. We created with wave equation modelling a large and diverse database of examples to feed to the network and showed that our algorithm could successfully transfer the knowledge it acquired from synthetic data to real data. The architecture of the network yields a pixel-wise classification enabling a high-resolution localization of the diffracting objects, while its probabilistic nature allows for some uncertainties in its answer. This method also lets us quickly scan through data migrated with different velocities to overcome the strong sensitivity of the diffraction images to velocity errors. Nevertheless, after carefully evaluating the results on field data, we found few false positives and false negatives and had difficulties to know how to better parameterize the algorithm to avoid those mistakes.

While deep neural networks can outperform every other method in classification tasks, they come with a number of disadvantages and difficulties. Deep learning is mostly empirical and works best when trained in a supervised fashion. It relies on the creation of a vast set of annotated data as well as on trials and errors to tune a large number of hyper-parameters. The datasets are usually prepared manually beforehand, and the performance is judged according to evaluation metrics computed on the training and validation sets. Those requirements are a challenge for applications in seismic interpretation. Because of inherent uncertainties in the data, the interpretation is often non-unique and subjective, and it also requires expertise in geology and geophysics. For this reason, as well as for the fact that most of the interpreted data is not publicly available, it is difficult to create large training databases. It also affects the reach of the evaluation metrics we use since they need to compare the answer of the network to non-perfect, and sometime non-existent, labels provided by interpreters.

To tackle the problem of creating a large labelled training dataset, we used synthetic modelling. This allows us to carefully control the subsurface model and provide an interpretation without manual work. However, synthetic data are a simplification of the reality and cannot account for all of the diversity and complexity that exists in nature. Other authors such as Serfaty *et al.* (2017) suggest to use a network pre-trained on publicly available datasets containing a very large number of annotated natural images. They then only need to label a small number of real seismic examples to fine-train the last layers of the network. This approach seems to work well but has few practical limitations. First, the geometry of the training data restricts the use of pre-trained nets to work on two-dimensional (2D) patches with three channels corresponding to the red, green and blue colour maps. This is a limiting factor for seismic data where structural objects are inherently three-dimensional and where full fold prestack data might bring more information as in this work or in automatic amplitude versus angle classification for instance. Furthermore, while it is understandable that shallow layers that learn to detect high-frequency characteristics such as edges are useful when transferred from natural images to seismic data, it is less intuitive for the deepest layers that have learnt abstract and large-scale concepts. Since the power of deep learning

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

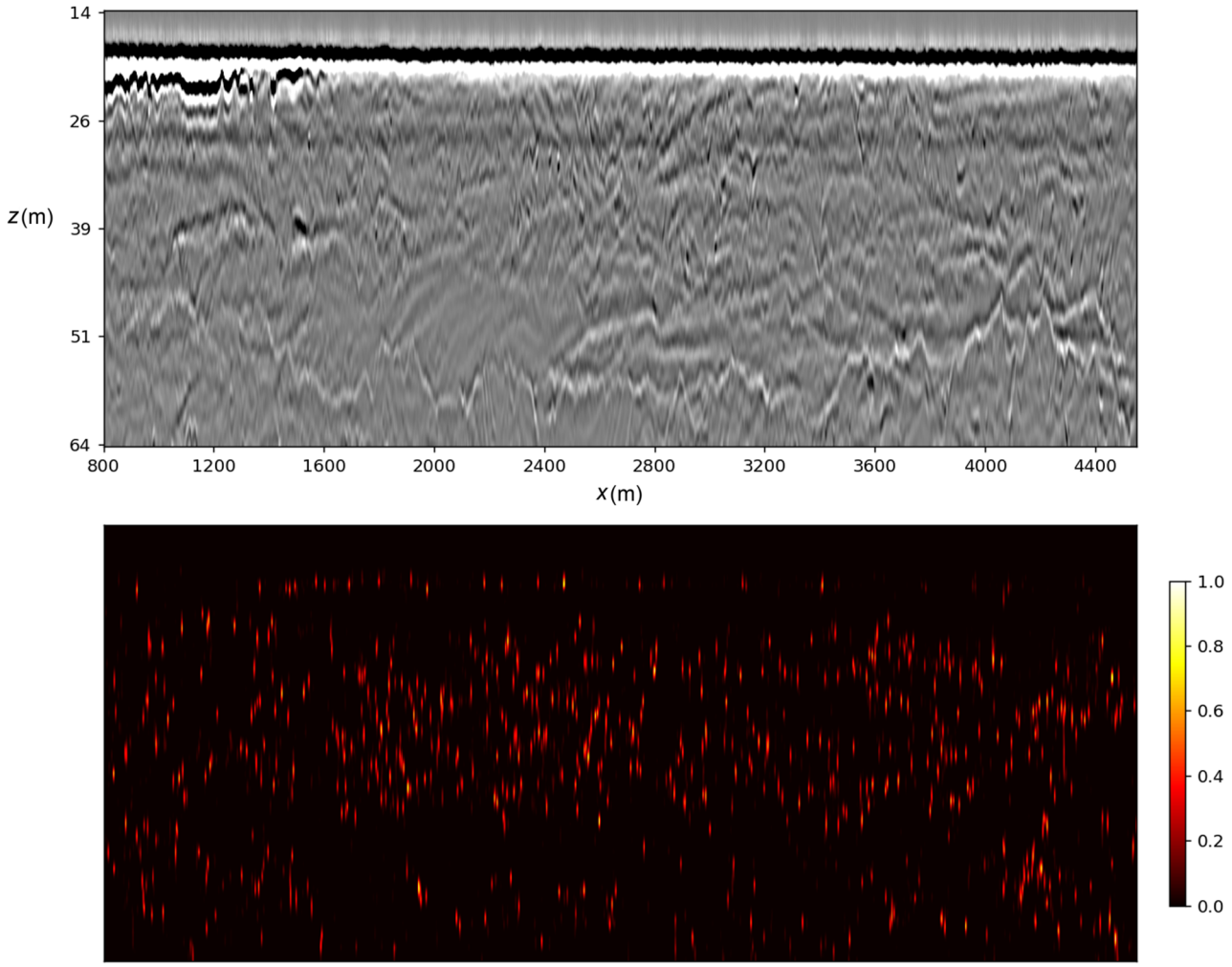

Figure 8 A full stack and the confidence attribute generated by the network. The local minima filter ran on the attribute counted 537 diffracting objects.

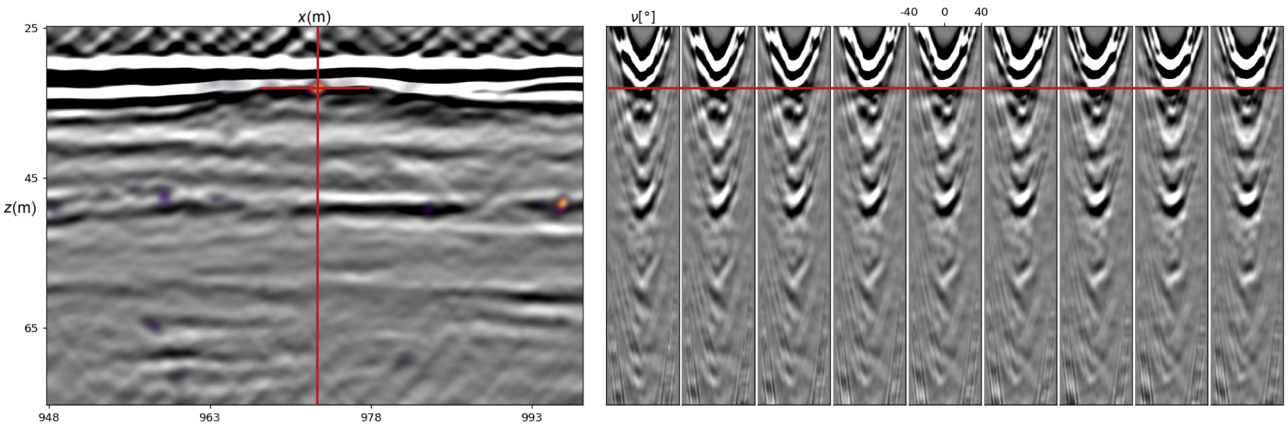

Figure 9 An example of true positive located just beneath the water bottom. The left image is a zoom on the stack with an overlay of the confidence attribute shown in Fig. 8. A vertical red line indicates the position of the central dip CIG displayed on the right side. Other gathers are displayed every 0.5 m around the central gather. The vertical red line highlights the signature of the boulder.

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

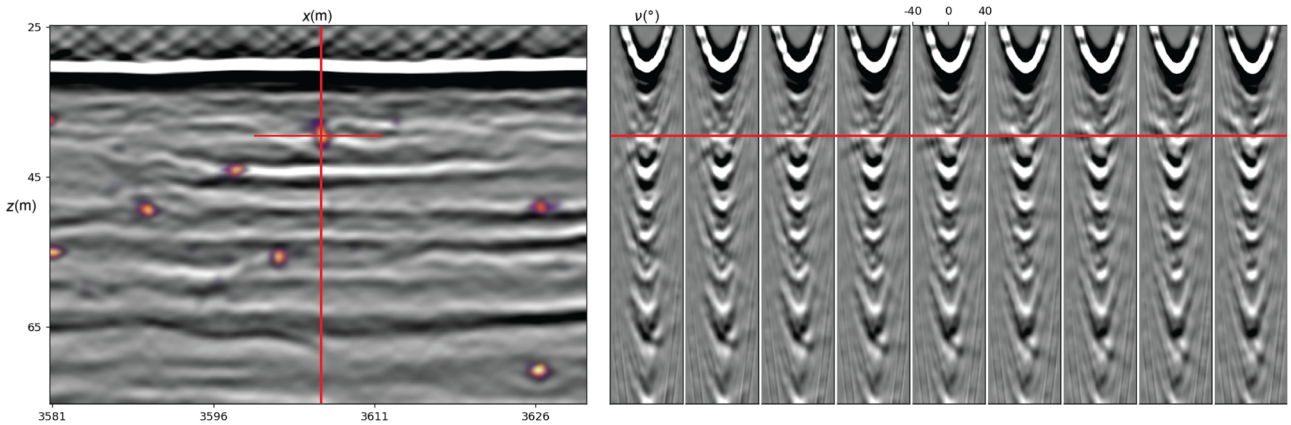

Figure 10 An example of true positive. Disposition of the figure is similar to Fig. 9.

comes from those abstract concepts learnt in the deeper layers of the network, it is unclear whether such pre-trained networks are really taking advantage of the full extent of deep learning and if traditional machine learning methods would not yield similar results. We believe that our approach using synthetic examples to train and progressively adding real examples as they are found by the network to fine train it is a good compromise to have full flexibility in the design in the absence of an ideal training database.

While it is common to judge the performance of a neural network with a numerical evaluation metric, we believe that such conclusion is more difficult to draw with geoscientific data. The labels we provide are often uncertain and sometimes wrong. In our work, for example, it is unrealistic to expect the network to know the exact position of the centre of the diffracting object as well as it exact spatial extent. Therefore, the minimum of the training loss function is probably not indicating the best possible parameterization of the network. Additionally, when evaluating our network on real data we observed a drop in performances compare to the blind evaluation on synthetic data. It is difficult to prove the generalization of the performances of a network. If the blind dataset we use for evaluation is too similar to the training data, a good evaluation score will not necessary extrapolate to all new data. Finally, we argue that providing an evaluation score itself is problematic when working with real data. Again, because of uncertainties and lack of perfect manual interpretations, it is not possible to know the *truth* and therefore to give a 0 to 1 score that is truly meaningful. We think that qualitative judgement by human experts of the machine's findings on field data, while subjective, is still required.

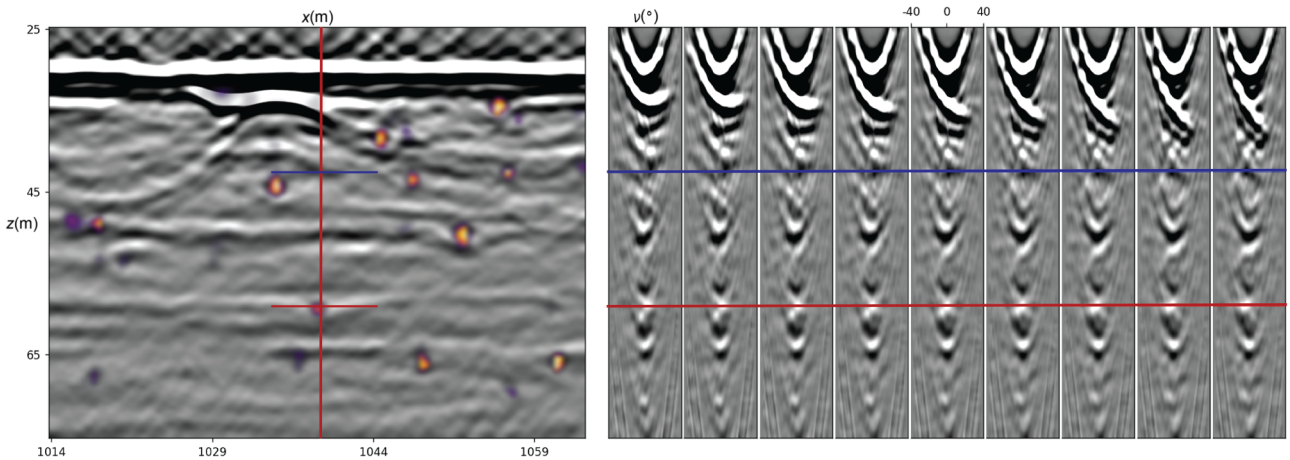

Figure 11 Examples of a false positive and a false negative. The disposition of the figure is similar to Fig. 9. The blue line indicates a probable diffraction that was not recognized by the network, while the red line highlights a cross-shaped signal wrongfully interpreted as a scatterer by the machine.

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers., *Geophysical Prospecting*, 1–15

In further work, we hope to benefit from the creation of a dataset of real examples by progressively incorporating real findings from our method on a variety of field-recorded data. We plan to extend the method to be three dimensional. In theory, such extensions do not pose any problem, the migration will yield three-dimensional (3D) spatial data with 2D azimuthal dip gathers. The CNN will be working in the same fashion but with 3D convolutions and a fourth dimension being channels made of the concatenation of azimuthal migration dip angles. We also plan to improve the precision of the method to perform automated residual velocity analysis, using the results of the network as a misfit criteria to optimize a tomographic inversion.

CONCLUSION

We have introduced a new method to automatically identify scattering points from prestack data migrated using diffraction imaging. We built a database of dip angle common image gathers containing point scatterers using wave equation modelling and trained a convolutional neural network to compute a spatially varying attribute, indicating the machine's confidence in having found diffracting objects in the subsurface. The use of synthetic data was a key in order to provide a variety of examples with their interpretation at minimal manual labour cost. We showed that our trained network could successfully transfer its knowledge on field-recorded data and bring a valuable help to interpreters on an engineering task. Additionally, this automated workflow enables us to quickly scan through different versions of the same dataset to account for potential errors in the migration velocities.

We also discussed some of the challenges associated with the use of artificial intelligence–based algorithms to analyse seismic data. Uncertainties and non-uniqueness of the interpretation as well as the non-guaranty of generalization of the results should be taken into account when evaluating the performance of a network. In particular, we believe that careful inspection by experts, while subjective and qualitative, should be nevertheless carried on a reasonable variety of real datasets before concluding that the problem at hand was solved. We see a good potential in our workflow and hope to prove it valuable in further work for applications to other structural and amplitude related interpretation tasks.

ACKNOWLEDGEMENTS

This project was partially funded by the Federal Ministry for Economic Affairs and Energy of Germany, project 832642. We thank our colleagues from Fraunhofer IWES for preprocessing the marine data used in this study.

DATA AVAILABILITY STATEMENT

Synthetic data: Data available on request from the authors. Real data: Author elects to not share data.

ORCID

Valentin Tschannen
https://orcid.org/0000-0003-1845-5487

REFERENCES

Arora Y. and Tsvankin I. 2017. Analysis of diffractions in dip-angle gathers for transversely isotropic media. SEG Technical Program Expanded Abstracts 2017, 1011–1016.

Audebert F., Froidevaux P., Rakotoarisoa H. and SvayLucas J. 2005. Insights into migration in the angle domain. SEG Technical Program Expanded Abstracts 2002, 1188–1191.

Cybenko G. 1989. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* **2**, 303–314.

Dafni R. and Symes W.W. 2017. Diffraction imaging by prestack reverse-time migration in the dip-angle domain. *Geophysical Prospecting* **65**, 295–316.

de Figueiredo J.J.S., Oliveira F., Esmi E., Freitas L., Schleicher J., Novais A., et al. 2013. Automatic detection and imaging of diffraction points using pattern recognition. *Geophysical Prospecting* **61**, 368–379.

Fomel S., Landa E. and Taner M.T. 2007. Poststack velocity analysis by separation and imaging of seismic diffractions. *Geophysics* **72**, U89–U94.

Guitton A. 2018. 3D convolutional neural networks for fault interpretation. Presented at the 80th EAGE Conference and Exhibition 2018.

Huang Y., Zhang D. and Schuster G.T. 2015. Tomographic resolution limits for diffraction imaging. *Interpretation* **3**, SF15–SF20.

Kanasewich E.R. and Phadke S.M. 1988. Imaging discontinuities on seismic sections. *Geophysics* **53**, 334–345.

Khaidukov V., Landa E. and Moser T.J. 2004. Diffraction imaging by focusing defocusing: an outlook on seismic superresolution. *Geophysics* **69**, 1478–1490.

Klokov A., Baina R. and Landa E. 2010. Separation and imaging of seismic diffractions in dip angle domain. 72nd EAGE Conference and Exhibition, Extended Abstracts.

Klokov A. and Fomel S. 2012. Separation and imaging of seismic diffractions using migrated dip-angle gathers. *Geophysics* **77**, S131–S143.

Landa E., Fomel S. and Reshef M. 2008. Separation, imaging, and velocity analysis of seismic diffractions using migrated dipangle

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers.,

gathers. SEG Technical Program Expanded Abstracts 2008, 2176–2180.

Landa E., Shtivelman V. and Gelchinsky B. 1987. A method for detection of diffracted waves on common-offset sections. *Geophysical Prospecting* **35**, 359–373.

LeCun Y., Bengio Y. and Hinton G. 2015. Deep learning. *Nature* **521**, 436–444.

LeCun Y., Bottou L., Bengio Y. and Haffner P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* **86**, 2278–2324.

Long J., Shelhamer E. and Darrell T. 2014. Fully convolutional networks for semantic segmentation. *CoRR* abs/1411.4038.

Merten D. and Ettrich N. 2015. GRT angle migration a 5D data mapping problem. 2015 International Conference on High Performance Computing Simulation (HPCS), 577–580.

Moser T. and Howard C. 2008. Diffraction imaging in depth. *Geophysical Prospecting* **56**, 627–641.

Pham N., Fomel S. and Dunlap D. 2018. Automatic channel detection using deep learning. SEG Technical Program, Expanded Abstracts, 2026–2030.

Reshef M. and Landa E. 2009. Post-stack velocity analysis in the dip-angle domain using diffractions. *Geophysical Prospecting* **57**, 811–821.

Ronneberger O., Fischer P. and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation. arXiv e-prints, arXiv:1505.04597.

Rumelhart D.E., Hinton G.E. and Williams R.J. 1988. Learning representations by back-propagating errors. *Cognitive Modeling* **5**, 1.

Sava P., Biondi B. and tgen J. 2005. Diffraction focusing migration velocity analysis. SEG Technical Program Expanded Abstracts 2004, 2395–2398.

Serfaty Y., Itan L., Chase D. and Koren Z. 2017. Wavefield separation via principle component analysis and deep learning in the local angle domain. SEG Technical Program Expanded Abstracts 2017, 991–995.

Shustak M. and Landa E. 2017. Time reversal based detection of subsurface scatterers. SEG Technical Program Expanded Abstracts 2017, 969–973.

Trorey A. 1970. A simple theory for seismic diffractions. *Geophysics* **35**, 762–784.

Waldeland A. and Solberg A. 2017. Salt classification using deep learning. Presented at the 79th EAGE Conference and Exhibition 2017.

Werbos P. 1974. *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. PhD thesis, Harvard University, Cambridge, MA.

© 2019 The Authors. *Geophysical Prospecting* published by John Wiley & Sons Ltd on behalf of European Association of Geoscientists & Engineers., *Geophysical Prospecting*, 1–15