DIFFRACTIVITY ATTRIBUTES: A CATEGORIZATION AND CORRELATING WITH SEISMIC REFLECTION ATTRIBUTES BY AUTOENCODER NETWORKS

S. Dell, J. Walda, A. Hoelker, and D. Gajewski

email: *sergius.dell@uni-hamburg.de*
keywords: *Diffractions, seismic attributes, deep learning*

ABSTRACT

Seismic attributes play a crucial role in delineating fault planes and mapping fracture density. Conventionally, seismic attributes derived from the migrated reflections are used for this purpose. The attributes derived from the other counterparts of the recorded wavefield are usually ignored since they have not been clearly categorized. We perform a categorization of the attributes derived from the diffracted part of the wavefield and combine them into a new seismic attribute class which we call diffractivity class. We distinguish three major sub-classes in the diffractivity class which describe geometric and amplitude properties of the seismic diffractions. To the geometric subclass, we assign point and edge diffraction focusing as well as the azimuth. The instantaneous amplitudes of the focused diffraction, which are isolated previously by a wavefield separation, belong to the amplitude subclass. We also group spectral decomposed diffraction amplitudes into a spectral decomposition class. The extraction of diffractivity attributes is based on 3D Kirchhoff time migration. Coherence along migration operators is first evaluated for different azimuth sectors. This provides point and edge diffraction focusing, and azimuth orientation in strike direction. Further, the reflection response is identified and attenuated by a dynamic operator muting. This yields focused diffraction amplitudes. The amplitudes are used to extract the instantaneous amplitudes attributes based on the complex-trace approach. The focused diffractions are subsequently spectral-decomposed using the wavelet-decomposition approach. We also propose to correlate the new diffractivity class with the conventional seismic reflection attributes. For classification and correlation of diffractivity attributes, we use a deep learning approach based on convolutional neural networks. The proposed method was successfully demonstrated on a 3D land data example acquired in the northern Switzerland.

INTRODUCTION

Seismic attributes represent quantitative measurements of the temporal or lateral characteristics of the seismic signal in the data (Brown, 2001; Chopra and Marfurt, 2005). The attributes are usually extracted from the migrated reflections, which implies that only the reflected part of the recorded wavefield is considered. The reflection attributes are well studied and usually grouped into classes according to the geometric, kinematic, dynamic, or statistical properties of the reflections. Roden et al. (2015) provided a categorization of seismic attributes based on six classes: instantaneous attributes, geometric attributes, amplitude accentuating attributes, AVO attributes, seismic inversion attribute, and spectral decomposition, which we also will use throughout the paper. However, the attributes extracted from non-reflected parts of the seismic data have been ignored in the most categorization. Recently, several attempts have been made to use a particular non-reflected part of the wavefield, seismic diffractions, to complement interpretation of reflection attributes. Alonaizi et al. (2013); Khoshnavaz et al. (2016); Ahmed et al. (2019) used coherence extracted along diffraction traveltimes as an attribute for a better structural interpretation particularly in hard-rock environments. Schoepp et al. (2015); Klokov et al. (2015) proposed mapping the diffraction energy at the reservoir level and correlating it with rates of initial production of the wells. Qualitative interpretation of the diffraction amplitudes was used for distinguishing edge-type and line-type diffractions, indicative of fault versus channel-fill features by (Burnett et al., 2015) and identifying an increased natural fracture density by Sturzu et al. (2015). The analysis of fracture network geometry from the composite diffraction signatures was studies by Grasmueck et al. (2015). In this paper, we generalize seismic diffraction attributes into a new class, diffractivity attributes. We also show how to correlate diffractivity attributes with the conventional reflection attributes. The proposed workflow is schematically displayed in Figure 1.

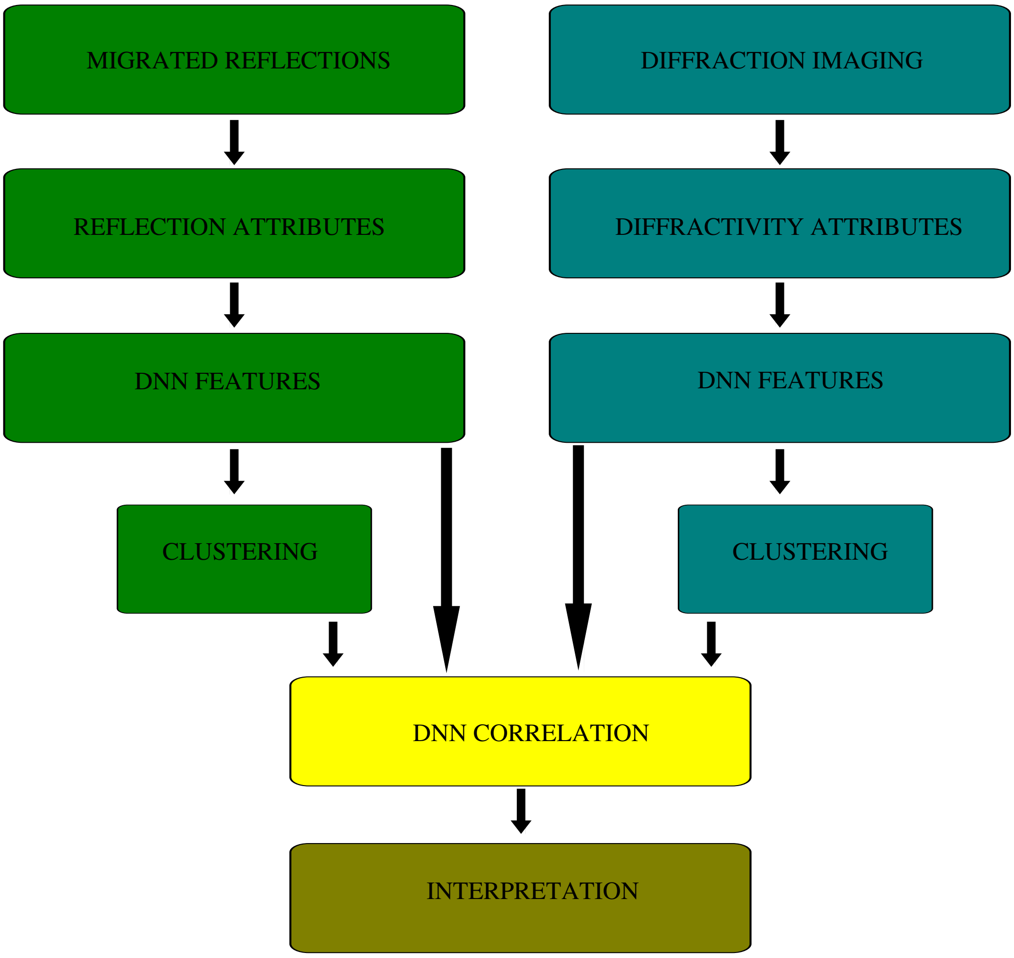


Figure 1: Figure shows a sketch of the entire workflow for a categorization of diffractivity attributes and their correlation with seismic reflection attributes by deep neural networks (DNN). Diffraction imaging and reflection migration are performed first. Diffraction imaging provides diffractivity attributes whereas the migration reflections are used to extract the seismic reflection attributes. Both classes of attributes are fed into separate autoencoder networks of the same structure, in order to extract the underlying features. These features are then combined on the top of the Siamese network for a DNN correlation. At the final step, correlated attributes are used for interpretation.

Diffracted waves describe backscattering from discontinuities in subsurface interfaces or structural elements much smaller than the prevailing wavelength of the incident seismic wave (Klem-Musatov, 1994; Kravtsov and Ning, 2010). We classify the diffracted waves of the first kind as the edge diffractions. The diffracted waves of the second kind are denoted as the point diffractions (e.g. Dell et al., 2019). A point diffractor equivalently scatters in all directions, while an edge diffractor (a fault, pinched-out layer, and so one) has an azimuthal directivity in its response. In 3D layered heterogeneous media, peak amplitude of edge diffractions will occur at an azimuth perpendicular to the strike direction of the fault. Moreover, seismic diffractions from edges have a phase reversal that occurs at the surface location just above the edge. The kinematic and dynamic differences in diffraction response allows one for using a modified Kirchhoff prestack time migration to isolate and subsequently focus point and edge diffractions. The modified Kirchhoff migration can also be used to evaluate semblance norm. The applying modified Kirchhoff operator provides not only volumes of focused diffraction amplitudes, but also azimuthal orientation of faults and coherence measure along the edge-tuned and point-tuned migration operators. The focused diffraction amplitudes, azimuthal orientation and semblances are used to build the diffractivity-attribute class.

Many methods have been developed to analyze different combinations of seismic attributes, e.g., principal component analysis in a combination with machine learning Roden et al. (2015). Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of correlated observations into a set of linearly uncorrelated variables called principal components. It transforms (decorrelates) a set of linearly related seismic attributes into a new set of independent attributes (see e.g., Duda et al., 2001). PCA has proven to be a robust method to reduce a large set of seismic attributes, to indicate variations in the data and to determine the most important attributes (e.g., Roden et al., 2015). However, PCA is prone to non-linear dependency as it is mainly focused on finding hidden linear correlations between seismic attributes. To better acknowledge non-linear dependencies in seismic attributes, several attempts have been made to complement PCA with unsupervised neural networks, e.g., self-organizing-maps (Roden et al., 2015). In this paper, we mainly exploit a strategy proposed by Walda et al. (2019). The authors describe a methodology to extract a set of uncorrelated attributes, so called feature cube, from subsets of seismic attributes. The method combines traditional machine learning and deep learning techniques and comprises a sample by sample automatic identification of spatial features. The authors used a deep convolutional autoencoder network (see e.g. Goodfellow et al., 2016). The network utilizes nonlinear encoder and nonlinear decoder functions, thus can learn a nonlinear generalization of standard PCA. Technically, the linear PCA can be implemented as an autoencoder with one hidden layer and without activation function. The encoder is trained to find a data representation (code) for an attribute image (input) which could be a 2D time-slice or 3D cube. The found code is used by a decoder for reconstructing the original image. With sufficient training, i.e., if the reconstructed image matches the input image sufficiently well, the autoencoder learns to capture the most salient features of the input image. The features are considered as salient if they exhibit significant changes in the brightness of a large number of pixels. These features are classified using an unsupervised clustering algorithm, the k-means algorithm, which results in a volume of features, so-called feature cube. The amount of clusters obtained through k-means can be varied, if other information is of interest. We also demonstrate how conventional reflection attributes can be correlated with the diffractivity class by a deep neural network (DNN). We suggest to use the Siamese Neural Network (SNN) (Bromley et al., 1993; Rao et al., 2016) as DNN correlation tool, whose architecture is capable of learning similarities between data sets. We also include a pre-training in design of our SNN network, i.e., we first use the above-described autoencoder method to extract and labels the features for geometric reflection attributes and diffractivity attributes, respectively.

This paper is organized as follows. In the next section, we introduce the concept of the diffractivity attributes and show how to extract them from the seismic data. In the subsequent section, we discuss a simplified categorization of the diffractivity attributes beginning with a short review of the network architectures used in our work. In the following section, we explain how to correlate new diffractivity attributes and the conventional reflection attributes. Finally, we demonstrate the presented approach on a 3D land data example.

0.1 DIFFRACTIVITY ATTRIBUTES

In this section, we describe how to categorize the diffractivity attributes and extract them from the seismic data. We begin with the outline of Kirchhoff time migration used for focusing seismic diffractions. Then we briefly summarize a deep learning approach we used to categorize diffractivity attributes and follow up with a more detailed introduction into classification of diffractivity attributes.

Extracting diffractivity from seismic data

The kernel of Kirchhoff migration in reflection imaging is a data summation along diffraction hyperbolas (an alternative approach comprising data spreading along diffraction hyperbolas can also be used). Thus, Kirchhoff migration per se is a diffraction imaging tool. A modification of the data summation tunes the Kirchhoff migration from a reflection towards a diffraction imaging tool. Considering a single image point (defined by X and Y position and a traveltime t the Kirchhoff operator defines a Δt surface, where Δt is a function of X/Y displacement from the image point. The X/Y displacements can be expressed as displacement azimuth and displacement distance. A reflection response occurs as a local spot on the operator surface (blue spot in Figure 2). The displacement azimuth relates to the dip direction of the reflector and the displacement distance relates to the dip angle. In reflection imaging all data on the operator are summed up. Constructive interference occurs at the reflection spot, while all other data cancel each other due to destructive interference. This means, the reflection should have a poor coherency over the entire migration operator as it occurs only in a local subset of the operator surface, in the vicinity of the stationary point. A point diffraction occurs as a diffuse signal all over the operator (grayish surface in Figure 2). Compared to a reflection response it has a weaker amplitude. However, a coherency measure along the entire operator surface provides a high semblance value due to constructive interference revealing the point-diffraction response. An edge diffraction can be seen as a mix of a reflection and point diffraction. On the migration operator surface it should show a response within an azimuth sector. The edge diffraction functions as a wavefield smoother on sharp reflector discontinuities. Due to the wavefield continuation, edge diffraction undergoes a phase reversal directly over the discontinuity and has a response signature (Klem-Musatov, 1994; Dell et al., 2019). This allows for measuring coherence tuned to the edge and point diffractions depending on the migration operator shape, i.e., grayish or orange surface. The coherence analysis provides the first diffractivity attributes: a focusing measure for point diffractions, a focusing measure for edge diffractions, and the azimuth. The azimuth attribute determines the local orientation of a fault to the acquisition line (Dell et al., 2019).

The weak signal amplitudes of diffractions and reflection left-overs quest for using a coherence measure, which generally provides a good and robust means for diffraction imaging. In order to directly work with focused amplitudes, we modify 3D Kirchhoff time migration. As the reflection response occurs confined in a vicinity of the stationary point (blue spot in Figure 1), its locality allows for muting specular energy Moser and Howard (2008). Dynamic operator muting, which we used to extract the amplitudes of focused diffractions, is implemented as followed: after Kirchhoff-moveout correction, the most energetic amplitudes are identified per time step, considered flatly as reflections and then muted. The remaining amplitudes are summed up. These focused amplitudes are used to generate the next two attributes classes: instantaneous amplitudes and spectral-decomposition.

Instantaneous seismic attributes describe the waveform shape (e.g., Taner et al., 1979; White, 1991) and are based on the complex-trace analysis, which treats the seismic trace as a real part of an analytic signal

$$x(t) = Re\left(A(t)e^{i\phi(t)}\right) , \tag{1}$$

where $x(t)$ is the recorded seismic trace, $A(t)$ is the instantaneous amplitude or reflection strength, $\phi(t)$ is the instantaneous phase. We also compute for our study the instantaneous frequency, F_I, and sweetness, S_I, given as

$$\begin{aligned} F_I &= \frac{1}{2\pi}\frac{d\phi(t)}{dt} , \\ S_I &= \frac{A(t)}{\sqrt{F_I(t)}} . \end{aligned} \tag{2}$$

We extract instantaneous amplitude, instantaneous frequency, instantaneous phase, and sweetness from the focused diffraction amplitudes. These attributes build a class of instantaneous amplitude attribute. We consider sweetness as an instantaneous attribute since it is calculated by dividing the instantaneous amplitude (reflection strength) by the square root of the instantaneous frequency. Note, some authors categorize sweetness as amplitude accentuating attribute (e.g., Roden et al., 2015). In interpreting reflection attributes, co-rendering sweetness with the discontinuity improves interpretation of coherency cubes as it reduces the

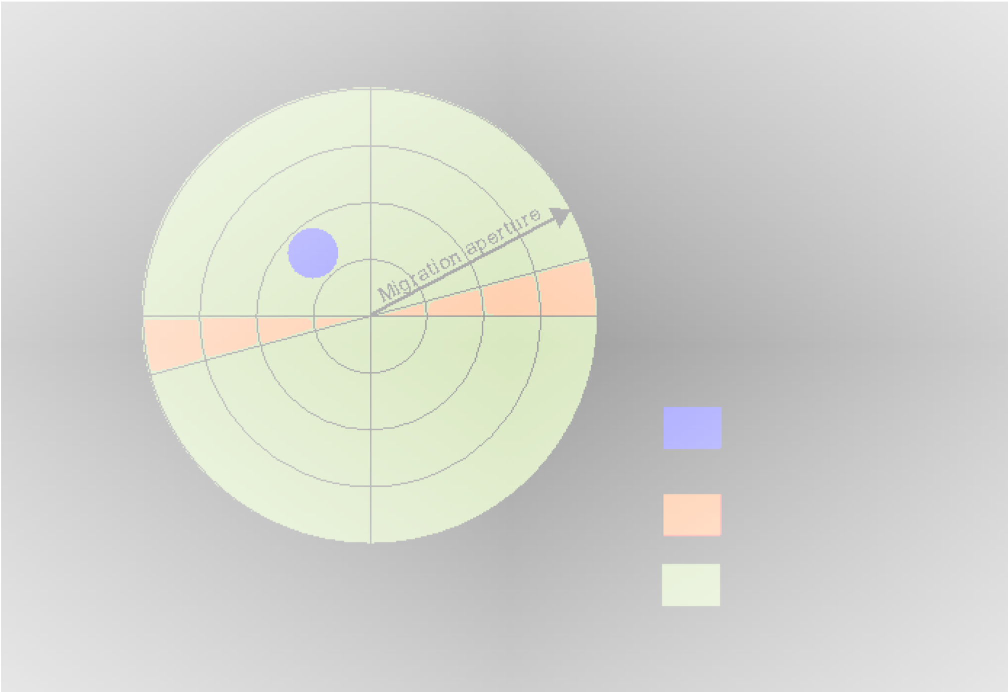


Figure 2: Figure shows a simplified sketch of Kirchhoff migration operator for a single image trace at an image time. Reflection imaging: the constructive sum of amplitudes from the blue spot will dominate the output while the destructive sum from the entire operator will cancel itself. Diffraction imaging: diffraction occurs as a diffuse signal either all over the operator (grayish surface) for point diffraction or over an operator sector (orange surface) for edge diffraction. Summing up along the entire operator or its sector is coherent.

contribution of high frequency events. Areas containing higher amplitudes and lower frequencies, e.g., sandy intervals, will show the highest values for sweetness, while lower amplitude and higher frequency sediments, e.g., thinly bedded shales, will display lower values for sweetness (Hart, 2008).

Finally, we decompose the focused diffraction amplitudes into different iso-frequency slices using wavelet transform (e.g., Meyer, 1992). The purpose of the attribute analysis based on the spectral decomposition is to compare the response of edges and small scatterer for various iso-frequencies.

In the next section, we show how to categorize diffractivity attributes based on feature maps (codes) learned by a convolutional neuronal network. We begin with a short description of a encoder-decoder network, so-called autoencoder, as this approach builds the skeleton of the proposed diffraction attribute classification and correlation with reflection attributes.

Classification of the diffractivity attributes

The classification of diffraction attributes is based on an autoencoder approach. An autoencoder is a neural network that is trained to attempt to copy its input to its output. The network comprises an encoder function $\mathbf{h} = \mathbf{f}(\mathbf{x})$ that encodes the input image, and a decoder function that produces a reconstruction of the input image $\mathbf{r} = \mathbf{g}(\mathbf{h})$ by means of the estimated code (weights). Internally, several hidden layers $\mathbf{h}$ of the encoder, using several weights matrices, are used to encode the input image. Autoencoder are constructed in a way which forces the weights to capture information only about that structure of the data-generating distribution which is needed to reconstruct training images (e.g. Goodfellow et al., 2016). This allows one to efficiently extract the most salient features. Autoencoders as well as linear factor models including principle-component-analysis (PCA) belong intrinsically to manifold learning methods (Hinton et al., 1997), which assume data concentrate around a low-dimensional manifold or a small set of such manifolds. The ability to learn the most salient features in a space of multidimensional seismic attributes make the autoencoder network particularly suitable for classifying the seismic attributes.

By implementation, an encoder is first trained to find a code (data representation) for an input image, which is a 3D subset around a sample of an attribute cube in our case. At the second step, a decoder uses the found code for reconstructing the original 3D subset. If the reconstruction of the input data lies as close as possible to the input data, the autoencoder captures the manifold structure of the data (e.g. Goodfellow et al., 2016), i.e., it learned the most salient features of the input image. Though we can use a loss function to estimate the error, it is not straightforward to estimate when an autoencoder has trained enough or is susceptible to overfitting. Seismic data in particular leads to a large variance in the loss function. Therefore, we observe reconstructions to estimate whether the reconstruction lies close to the input. This is a human guided quality control step in an otherwise fully automatic process. Figure 3 shows an example of the autoencoder network used in this paper. In the first half, the data is encoded into a vector of length 128. These 128 codes contain the most important features of the considered attributes and there spatial behaviour. The second half decodes the encoding to the original attribute sections. The input of the first layer and output of the last layer are compared by a misfit function, in our case mean squared error. The filters are updated reducing the misfit similar to conventional inversions.

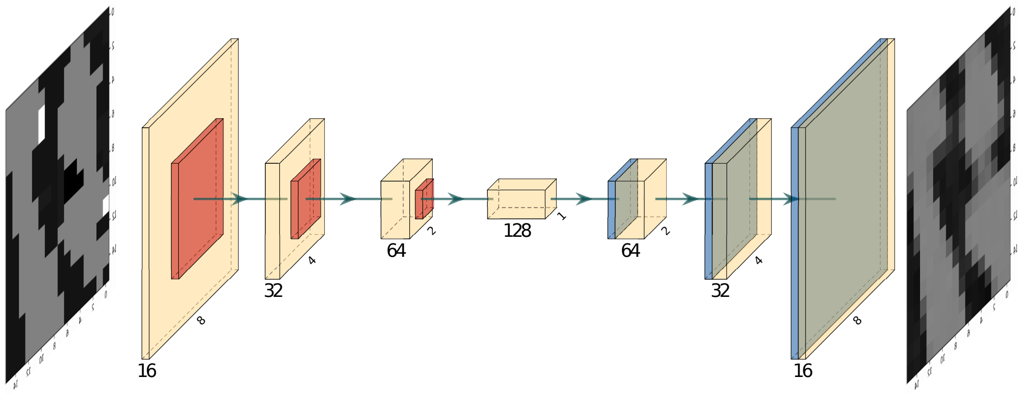


Figure 3: Autoencoder network (after Walda et al. (2019)). Structure of the autoencoder network used in this work. In the first half, the data is encoded into a vector of length 128. These 128 codes contain the most important features of the considered attributes and there spatial behaviour. The second half decodes the encoding to the original attribute sections. The input of the first layer and output of the last layer are compared by a misfit function, in our case mean squared error. The filters are updated reducing the misfit similar to conventional inversions. Though the network shows a timeslice example for visualization purposes, the actual data used is 3D.

We distinguish three major sub-classes in the diffractivity attributes. The first subclass is built based on kinematic properties of the diffractions. We categorize this subclass as the geometric attributes. This includes the point and edge diffraction focusing as well as the azimuth. Instantaneous amplitude, phase and frequency, and sweetness build the second subclass, instantaneous diffraction amplitude attribute. The third subclass consists of focused diffractions decomposed into iso-frequencies. In this paper however, we use geometrical and instantaneous amplitude classes only to classify diffraction. Using the spectral decomposition class requires designing a recurrent neural network (RNN) whereas the frequency axes is considered as the time axes. The purpose of the attribute analysis based on the spectral decomposition is to compare the response of edges and small heterogeneities for various iso-frequencies. Particularly, iso-frequency volumes can be used to train a convolutional network to recognize patterns of focused edge and point diffractors. Since the edge diffractor should scatter frequency-independent, all features should cluster in the same location, while this is not the case for the point diffractor. Therefore, the network will be built to classify whether a focused event is an edge or point diffractor. RNN will consume the sequence of iso-frequency volumes as a time-based sequence, one item per time step of the RNN.

Our categorization method also comprises traditional machine learning for feature classification. The features (codes) are classified using an unsupervised clustering algorithm, the k-means algorithm (MacQueen et al., 1967). The purpose of k-means algorithm is to group similar features together by discover underlying patterns. To achieve this objective, k-means searches for a fixed number of clusters in the feature space. A cluster refers to a collection of feature points aggregated together because of certain similarities. The center of the cluster (its location) is defined as a centroid. The number of centroids is usually determined by the user as a target number k. Every feature point is allocated to each of the clusters by minimizing mean-squared distances. In deep, the k-means algorithm initializes a first group of randomly selected centroids. These centroids are used as the starting points for every cluster. The positions of the centroids are iteratively optimized while keeping the centroids as small as possible. The k-means algorithm converges if the centroids have stabilized, i.e., there are no changes in their positions. The result is a so-called feature cube, ideally with only three clusters corresponding to the edge diffractions, point diffractions and other undesired events (so-called garbage collection). Considering the noise and reflection left-overs as information carrier, we chose in our work six clusters, $k = 6$, as the target number (Walda et al., 2019).

In the next section, we show how to correlate diffractivity attributes with reflection attributes. We begin the section with a brief description of a particular CNN design, so-called Siamese CNN.

Correlating diffractivity and seismic reflection attributes

Reflection attributes aims at revealing trends or patterns in seismic data and predicting a seismic facies or a property such as porosity Chopra and Marfurt (2005). Our main objective is delineation of faults and fractures. Thus, we consider those particular attributes which describe, or can be co-rendered to, the lateral continuity of the seismic signal. Lateral properties of reflections are best described by the geometric attributes. In this work we use discontinuity, dip, azimuth, mean curvature, the most negative and positive curvature (Walda et al., 2019). Our aim is mapping two different attribute classes through a network into a 3D similarity space. As we deal with multidimensional attribute classes, we also aim to learn the low dimensional (underlying) structure of attributes without explicit labeling input data. Siamese neural networks fit these modeling requirements perfectly Rao et al. (2016). Siamese neural networks are comprised of two neural networks that take a pair of images as input and share a common loss function (Chopra et al., 2005). The two networks of a SNN are identical to each other in their architecture, and they share the same weights. In our work, we use pre-trained CNN features in the design of our siamese neural networks. We obtained the pre-trained features by applying the autoencoder networks as described above, to reflection attributes and diffractivity attributes individually. The pre-trained features (labels) are fed into two sub networks which generally have the same architectures and weights. They are combined in the end by a fully connected neural network that takes the output of both autoencoders and combines the found encodings to measured distances between the input features. Figure 4 shows the architecture of the SNN used in our work.

DATA EXAMPLE

We used a subset of 3D-seismic land data that Nagra acquired in northern Switzerland. The whole survey was recorded in winter 1997 as a 3D cross-spread acquisition (Birkhäuser et al., 2001). Line spacing was 180 m and station interval was 30 m for sources and receivers. The survey has 8906 source points, 9036 receiver stations and approximately 3.8 million traces. The recording time was 4s. Mixed source types, two vibrator sweeps and several dynamite configurations were used. The sweeps cover the bandwidth 10 to 100 Hz. The nominal fold of coverage is 20. The data were reprocessed in 2016 in a fairly conventional manner: deconvolution, de-noising, amplitude recovery, refraction static corrections and NMO stack (Hölker and Birkhäuser, 2018). Additionally, we applied a planar-event attenuation as described in Dell et al. (2019) to NMO stack, in order to obtain reflection-attenuated stacks. We used a subset with an inline range of 1408 - 1775, crossline range of 2200 - 2450, and time range of 0 - 1200 ms. The inline/crossline bin size is 15.00 m/line. The inline orientation is 30.64 degrees from Northing. The subset area is 14.32 squared km.

Our visualizations focus on a prominent dipping horizon at a depth range 800 – 1180m (referred to as Stubensandstein), taking advantage of preliminary interpretations by NAGRA (2018). The reflection-attenuated stacks still contain residual reflections (see e.g. Dell et al., 2019). This is relevant to diffraction

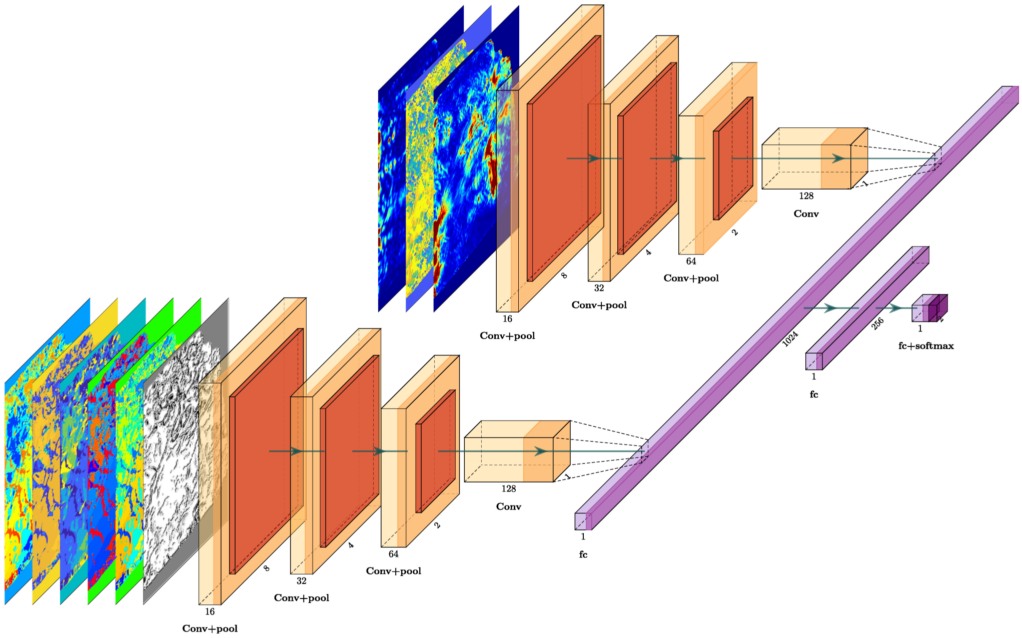


Figure 4: Model architecture of a siamese network. It contains two sub-networks with the same architecture (autoencoders). Encoded features from both sub-networks are combined at the top level of a fully connected neural network, which is trained to minimize distance between similar inputs.

imaging, because residual reflection energy may be locally stronger than the desired diffraction energy and thus obscuring the diffraction images. To mitigate this problem, we applied a dynamic operator muting to data to further eliminate residual reflection energy. We considered the data after the operator muting as focused diffractions and used them to extract the instantaneous amplitude attributes as well as for spectral decomposition. To produce geometric diffraction attributes, we used a coherence analysis in azimuth sectors. The coherence measure along an azimuth has proven to be very robust to reflection left-overs.

Figure 5 shows the geometrical attribute class: edge diffraction focusing, point diffraction focusing, and azimuth. Figure 6 displays the instantaneous amplitude class: instantaneous phase, instantaneous frequency, sweetness. Moreover, we provide the time-slice of the conventionally migrated reflection data (Figure 6f) and focused diffractions after dynamic operator muting (Figure 6e). The sections are extracted for the geological horizon Top Stubensandstein. We observe, the diffractivity attributes are rather randomly scattered in areas where geologically a homogeneous medium is expected. However, they exhibit high-coherent pattern in areas where the subsurface is geologically inhomogeneous, which is directly related to the increase in the intensity of the diffraction response. As shown in the reflection image (Figure 6f), a discontinuity is also assumable due to its amplitude pattern. Capturing the discontinuity however would require picking and distinguishing from other amplitude pattern, i.e., it requires a skilled interpretation. The images of diffractivity attributes show the discontinuity directly as isolated signal which may be useful as interpretation support or serve as a means to validate on existing interpretation or classify interpreted structural elements and is complementary to reflection imaging.

Figure 7 shows the spectral decomposition class. Iso-frequency slices range from 20 Hz to 106 Hz. Iso-frequency slices show the different frequency character for faults and point scatters. Several point scatterers are only visible for low frequencies while the discontinuity under investigation (inline 1600, crossline 2275) is present in all frequency slices.

In this paper, we used the geometrical and instantaneous amplitude attributes for classification. We did not use spectral decomposition attributes as it required a change in the design of the network, which we briefly discussed above. Figure 8 shows the result of encoding-decoding for a time slice at 600 ms for two attributes: edge diffraction focusing (right column) and edge diffraction azimuth (left column). Figure 8a,c,e) display the input image, reconstructed image, and the difference plot, respectively, for the edge

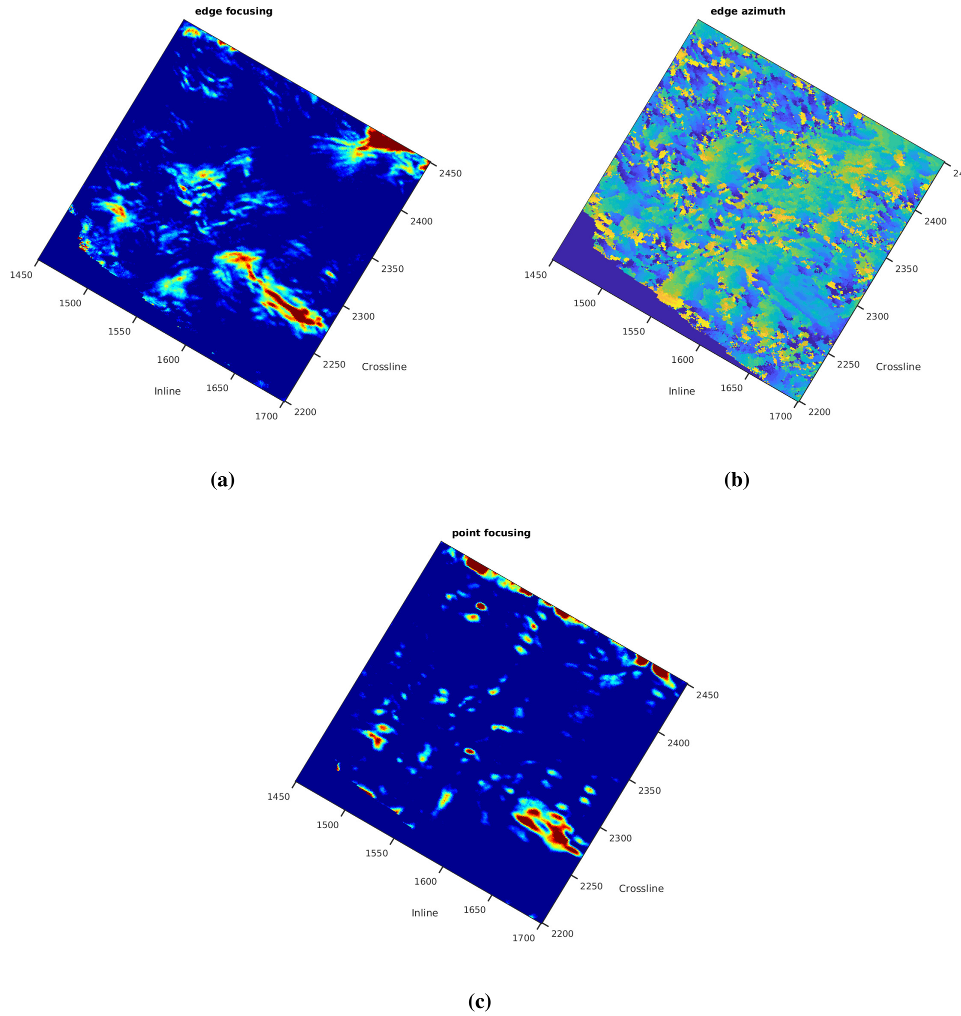


Figure 5: Geometrical attributes extracted for a geological horizon, Top Stubensandstein. a) Edge diffraction focusing, b) edge diffraction azimuth, and c) point diffraction focusing.

diffraction focusing. Figure 8b,d,f) display the input image, reconstructed image, and the difference plot for the edge diffraction azimuth. We observe a very good reconstruction of the input images. Note, the purpose of the autoencoder is not to perfectly reconstruct the image, i.e., to learn an identity function. The aim of the autoencoder networks is rather to learn the underlying representation (code) that generates data. This representation should capture the most salient features, which we also observe in the reconstructed images, which explains the difference plots.

For our autoencoder network, we used an architecture consisting of four hidden layers, each comprising a convolution layer, batch normalization and max pooling. The first convolutional layer filters the $16x16x16x6$ input image with 16 kernels of size $3x3x3$ with a stride of one. The second convolutional layer takes as input the (batch-normalized and pooled) output of the first convolutional layer and filters it with 32 kernels of size $3x3x3$ with a stride of one. The third convolutional layer has 64 kernels of size $3x3x3$ connected to the (normalized, pooled) outputs of the second convolutional layer. The last convolutional layer has 128 kernels of size $3x3x3$ connected to the (normalized, pooled) outputs of the third convolutional layer. After training the model with a base learning rate of $1e-4$ for first ten thousands iteration, we decrease it to $1e-5$ for the remaining training. Training the model on a single GPU, *GeForce GTX 1080*, took around three days. In total, three millions iteration were performed. The batch size was 256 by 16 by 3.

The encoded features were classified using an unsupervised clustering algorithm, the k-means algorithm, which results in a volume of only six clusters corresponding to edge diffraction, point diffractions and other events (a garbage collection). Figure 9a) displays a time slice of all labeled features for the geological horizon Top Stubensandstein. We interpreted the labels displayed in Figure 9c) as point diffractions indicating paleo-topography, and labels shown in 9b) as edge diffractions indicating faults and fracturing.

We used the model trained on diffractivity attributes for a correlation with reflection attributes. To build our SNN, we also trained a model using geometrical reflection attributes. Figure 10 shows geometrical reflection attributes: (a) discontinuity, (b) dip, (c) azimuth, and (d) the most-positive curvature. The attributes were directly extracted from the 3D time-migrated reflections by a local structure-oriented semblance analysis (Walda et al., 2019). We note, as semblance represents rather a very smooth function, we apply to semblance sections so-called local-contrast normalization technique (Brady and Field, 2000). Other attributes were prepared for training by the conventional standardization method (e.g., in Yu et al., 2006). An encoder was again trained to find a data representation (code) for a patch of reflection attributes. Then, a decoder used the found code for reconstructing the original attributes. In total, two millions iteration were performed as in the case of the diffractivity attributes. At the final step of pre-training the reflection model, the encoded features were classified using the k-means algorithm, which is shown in Figure 10 e and f.

We finally used reflection and diffraction attributes, i.e, their corresponding features, for a DNN correlation by a Siamese neural network. We used leaky ReLU (rectified linear units) activation functions throughout the architecture of SNN to avoid dead neurons during the training process. We set the base learning rate to $1e-4$. The fully connected layers have 1024, 256, and 6 neurons, respectively. Training the model on a single GPU took around 2 hours. The layers were initialized with Xavier initialization and stochastic batch gradient descent was used during training. Figure 11 shows the result of applying SNN to pre-trained models. In Figure 11 we display the DNN correlated features. Figure 11 a) show the features with the highest correlation weights which most likely correspond to the discontinuities. Figure 11 b) show the features with lower correlation weights.

Figure 11 a) also displays the interpreted discontinuities (red lines) overlaid with the highest correlation weights. We observe, the discontinuities are more continuous in the section comparing to the interpreted labels for reflection attributes (Figure 10f) or diffractivity attributes (Figures 9b). Particularly, the vanishing structure at IL 1650/ XL 2275 is clearly visible. It is worth to mention, some diffractivity attributes, e.g., edge focusing and focused diffraction amplitude, also reveal a structure at IL 1650/ XL 2275, The structure however appears rather disturbed than continuous. Nevertheless, diffractivity attributes indicate structures which are characteristic for discontinuities, so that the DNN correlation shows the high weights and make possible to capture a discontinuity at IL 1650/ XL 2275. We observe a complex geometry of discontinuity which might explain a fairly broad diffractivity response at this location and missing the structure in the reflection-based coherence section. We also note the weak diffractivity response for a continuous discontinuity crossing the whole section in the upper part. This could be explained by either the discontinuity represents rather a flexure than a fault or some diffractivity was attenuated by pre-processing, particularly NMO-stacking.

DISCUSSION

In this paper, we performed a rather simplified categorization of a complex diffraction phenomenon into point and edge diffractions. In the literature, there are other classifications of the diffraction phenomenon available (see e.g. in Moser, 2011). The presented method for characterization of diffractions however can be modified in order to account for other diffraction types. For instance, tip diffractions could be of interests for model building, imaging, and interpretation of such structures as pinch-out layers. Including tip diffractions in classification may require to swap the distance-based clustering (k-means algorithm) to a density-based clustering non-parametric algorithm, e.g., DBSCAN (Ester et al., 1996). For a set of points in some space, it groups together points that are closely packed together (points with many nearby neighbors), marking as outliers points that lie alone in low-density regions (whose nearest neighbors are too far away) (Ester et al., 1996). These two disadvantages of k-meansâĂŤits lack of flexibility in cluster shape and lack of probabilistic cluster assignmentâĂŤmean that for many datasets (especially low-dimensional datasets) it may not perform as well as one might hope.

We trained our network using the mean-squared-error loss function. The mean-squared error applied to the pixels of an image implicitly specifies that an underlying cause is only salient if it significantly changes the brightness of a large number of pixels (Goodfellow et al., 2016). This can be problematic if the task we wish to solve involves interacting with small objects like diffraction response form very small inhomogeneities. This issue usually arises if the capacity of the autoencoder is not sufficient, which means the training with the mean-squared error will not identify the small inhomogeneities as being salient enough to encode. To overcome the limited capacity issue we can add more data or increase the complexity of the model by e.g. using more hidden layers. Furthermore, another loss function such as mean-absolute-error might be better suited.

The used Siamese CNN topology is one of the described in the literature. Depending on the point where the information from each twin branch is combined, other architectures could be applicable (e.g. Leal-Taixé et al., 2016). Combining twin images at the top layers of each branch and a subsequent feeding the loss function aims at learning a manifold in a space, where different classes are easily separable, i.e., are highly correlated. In this case, input patches are processed by two parallel branches featuring the same network structure and weights. This type was very appealing for us as diffractivity and reflection attributes are mutually orthogonal due to the different physics of reflection and diffractions, but describes the same underground medium. Twin image patches could already be merged inside the network, i.e., during the feature learning. In this case, the top layers of the parallel branches processing the twin inputs are concatenated and some more layers are added on top of that. The clustering might be performed in an adaptive fashion (Chang et al., 2017). Finally, reflection and diffraction attributes can be used as a joint data input. Here, the two input patches are stacked together forming a unified input to the convolutional neural network. Jointly using information from twin branches from the first layer does not require pre-trained labels (Zagoruyko and Komodakis, 2015).

Processing the refection attributes revealed a structure, which was not observable in the coherence sections but visible in the labeled features. However, the structure was visible in the diffractivity attributes, which are the direct indicators for discontinuities, and also in the corresponding labeled features though slightly mispositioned. The DNN correlation of reflection and diffractivity features revealed a complex geometry of the presumable structure that may compose two parallel discontinuities. A careful evaluation of the results and including the present geological information is needed to make a final conclusion on the discontinuity type, i.e., flexure or fault. Moreover, the proposed method requires a diffraction-friendly processing of seismic data. Particularly NMO stacking is tuned to optimally stack the seismic reflections, which surely lead to a certain attenuation of diffractions before extracting diffractivity attributes. It is therefore advisable to directly extract diffractivity attributes from the prestack data to assure the maximum on diffraction information is fed into the deep neural network.

CONCLUSIONS

We presented a method to perform a classification of seismic diffractions based on the combination of machine and deep learning approaches. We first extracted the diffracted part of the wavefield by 3D Kirchhoff time migration using a dynamic reflection muting. On the fly we also evaluated coherence along Kirchhoff migration operators for different crossline-inline azimuths. In the end, we obtained focused diffraction amplitudes, coherence, and azimuths, which we then used to build a new class of the seismic attributes, which we call the diffractivity class.

We distinguished three major sub-classes in the diffractivity class which describe geometric and amplitude properties of the seismic diffractions. Point and edge diffraction focusing as well as the azimuth are assigned to the geometric subclass. The instantaneous amplitude attributes extracted by a conventional complex-trace analysis applied to the focused diffractions are considered as the amplitude subclass. Iso-frequency volumes obtained by the wavelet-decomposition approach are grouped into the spectral decomposition class. We used the geometric and instantaneous amplitude attributes for clustering by a combination of the convolutional autoencoder network and k-means classifier. The auotencoder was used for learning the underlying data representations, features that generate the diffractivity attributes, while k-means was used at the final step for clustering these features.

Furthermore, we showed how to correlate the new diffractivity class with the conventional seismic reflection attributes. For correlation of diffractivity attributes, we used features generated by the autoencoder network and siamese topology architecture of the convolutional networks to evaluate the similarity between diffraction and reflection features. The proposed method was successfully demonstrated on a 3D land data example acquired in the northern Switzerland.

REFERENCES

Ahmed, K., Dell, S., Schwarz, B., Gajewski, D., and Ettrich, N. (2019). 3-d seismic imaging in crystalline rock environments: An approach based on diffraction focusing. *Journal of Applied Geophysics*, 65:49 – 59.

Alonaizi, F., Pevzner, R., Bóna, A., Shulakova, V., and Gurevich, B. (2013). 3d diffraction imaging of linear features and its application to seismic monitoring. *Geophysical Prospecting*, 61(6):1206–1217.

Birkhäuser, P., Roth, P., Meier, B., and Naef, H. (2001). 3D Seismik: Räumliche Erkundung der mesozoischen Sedimentschichten im Zürcher Weinland. *Nagra technical report*, NTB 00-03:Nagra, CH–5430 Wettingen.

Brady, N. and Field, D. J. (2000). Local contrast in natural images: Normalisation and coding efficiency. *Perception*, 29(9):1041–1055.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1993). Signature verification using a "siamese" time delay neural network. In *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93, pages 737–744. Morgan Kaufmann Publishers Inc.

Brown, A. R. (2001). Understanding seismic attributes. *Geophysics*, 66(1):47–48.

Burnett, W. A., Klokov, A., Fomel, S., Bansal, R., Liu, E., and Jenkinson, T. (2015). Seismic diffraction interpretation at piceance creek. *Interpretation*, 3(1):SF1–SF14.

Chang, J., Wang, L., Meng, G., Xiang, S., and Pan, C. (2017). Deep adaptive image clustering. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5880–5888.

Chopra, S., Hadsell, R., and LeCun, Y. (2005). Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546.

Chopra, S. and Marfurt, K. J. (2005). Seismic attributes — A historical perspective. *Geophysics*, 70(5):3SO–28SO.

Dell, S., Hoelker, A., and Gajewski, D. (2019). Using seismic diffractions for assessment of tectonic overprint and fault interpretation. *Geophysics*, 84(1):IM1–IM8,doi.org/10.1190/geo2018–0113.1.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification*. Wiley Interscience, New York.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pages 226–231. AAAI Press.

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

Grasmueck, M., Moser, T. J., Pelissier, M. A., Pajchel, J., and Pomar, K. (2015). Diffraction signatures of fracture intersections. *Interpretation*, 3(1):SF55–SF68.

Hart, B. S. (2008). Channel detection in 3-d seismic data using sweetness. *AAPG Bulletin*, 92(6):733.

Hinton, G., Dayan, P., and Revow, M. (1997). Modeling the manifolds of images of handwritten digits. *IEEE Transactions on Neural Networks*, 8:65–74.

Hölker, A. and Birkhäuser, P. (2018). *3D-Seismik Datenbearbeitung 2016-2017*. Nagra Arbeitsbericht NAB 17-36, Wettingen, Switzerland.

Khoshnavaz, J., Bona, A., Hossain, M., Urosevic, M., and Chambers, K. (2016). Diffractivity – another attribute for the interpretation of seismic data in hard rock environment, a case study. *Interpretation*, 4:B23–B32.

Klem-Musatov, K. (1994). *Theory of seismic diffractions*. SEG.

Klokov, A., Irkabaev, D., Ogiesoba, O. C., and Munasypov, N. (2015). Correlation between seismic diffractions extracted from vertical seismic profiling data and borehole logging in a carbonate environment. *Interpretation*, 3(2):T121–T129.

Kravtsov, Y. A. and Ning, Yan., Z. (2010). *Theory of diffraction. Heuristic approaches*. Alpha Science.

Leal-Taixé, L., Canton-Ferrer, C., and Schindler, K. (2016). Learning by tracking: Siamese cnn for robust target association. *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 418–425.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Meyer, Y. (1992). *Wavelets and Operators*. Cambridge University Press.

Moser, T. J. (2011). Edge and Tip Diffraction Imaging in Three Dimensions. In *73th Conference and Technical Exhibition, EAGE*, page P148. Extended Abstracts.

Moser, T. J. and Howard, C. B. (2008). Diffraction imaging in depth. *Geophysical Prospecting*, 56:627–641.

NAGRA, editor (2018). *Preliminary horizon and structure mapping of the Nagra 3D seismics ZNO-97/16 (Zürich Nordost) in time domain*. Arbeitsbericht NAB 18-36, Wettingen, Switzerland.

Rao, S. J., Wang, Y., and Cottrell, G. W. (2016). A deep siamese neural network learns the human-perceived similarity structure of facial expressions without explicit categories. In *CogSci*.

Roden, R., Smith, T., and Sacrey, D. (2015). Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps. *Interpretation*, 3(4):SAE59–SAE83.

Schoepp, A., LabontÃl', S., and Landa, E. (2015). Multifocusing 3d diffraction imaging for detection of fractured zones in mudstone reservoirs: Case history. *Interpretation*, 3(1):SF31–SF42.

Sturzu, I., Popovici, A. M., Moser, T. J., and Sudhakar, S. (2015). Diffraction imaging in fractured carbonates and unconventional shales. *Interpretation*, 3(1):SF69–SF79.

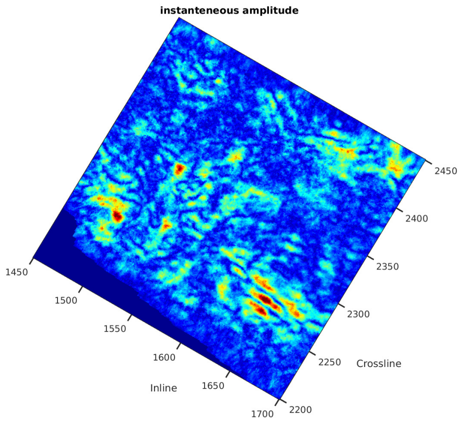
Taner, M. T., Koehler, F., and Sheriff, R. E. (1979). Complex seismic trace analysis. *Geophysics*, 44(6):1041.

Walda, J., Dell, S., and Gajewski, D. J. (2019). Unsupervised seismic attribute interpretation using deep learning. *Geophysics*, submitted.

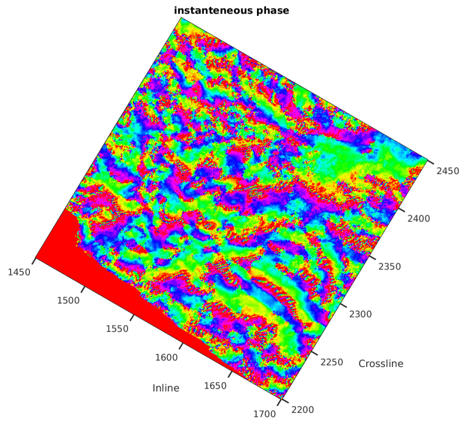
White, R. (1991). Properties of instantaneous seismic attributes. *The Leading Edge*, 10(7):26.

Yu, L., Wang, S., and Lai, K. K. (2006). An integrated data preparation scheme for neural network data analysis. *Knowledge and Data Engineering, IEEE Transactions on*, 18:217– 230.

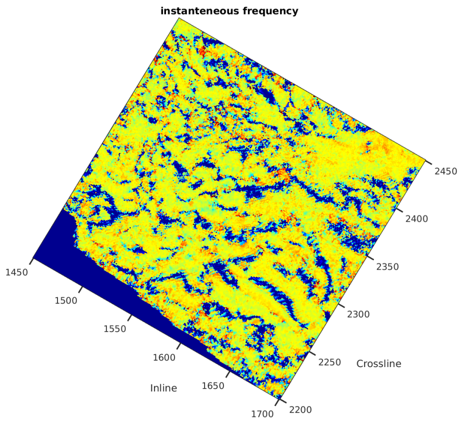
Zagoruyko, S. and Komodakis, N. (2015). Learning to compare image patches via convolutional neural networks. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4353–4361.



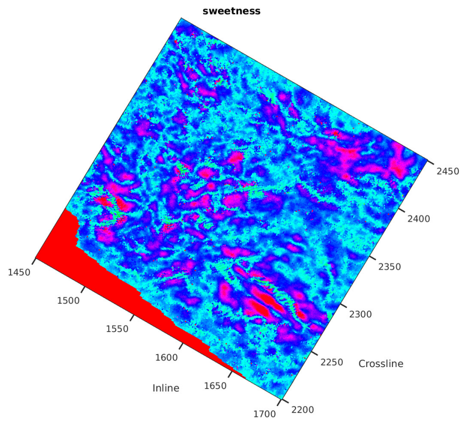
(a)



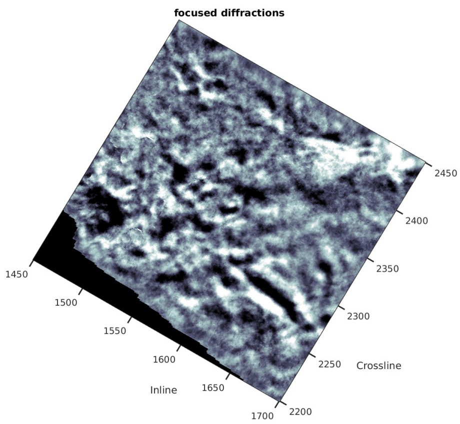
(b)



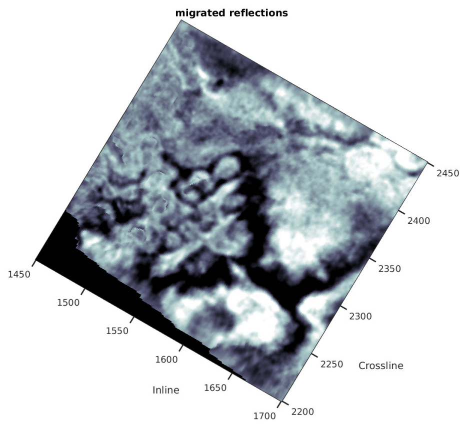
(c)



(d)



(e)



(f)

Figure 6: Instantaneous amplitude attributes extracted for a geological horizon, Top Stubensandstein. a) Instantaneous amplitude, b) instantaneous phase, c) instantaneous frequency, d) sweetness. The focused diffracted amplitudes used to extract the instantaneous amplitude attributes are shown in e). For comparison, the time-migrated reflections are displayed in f).

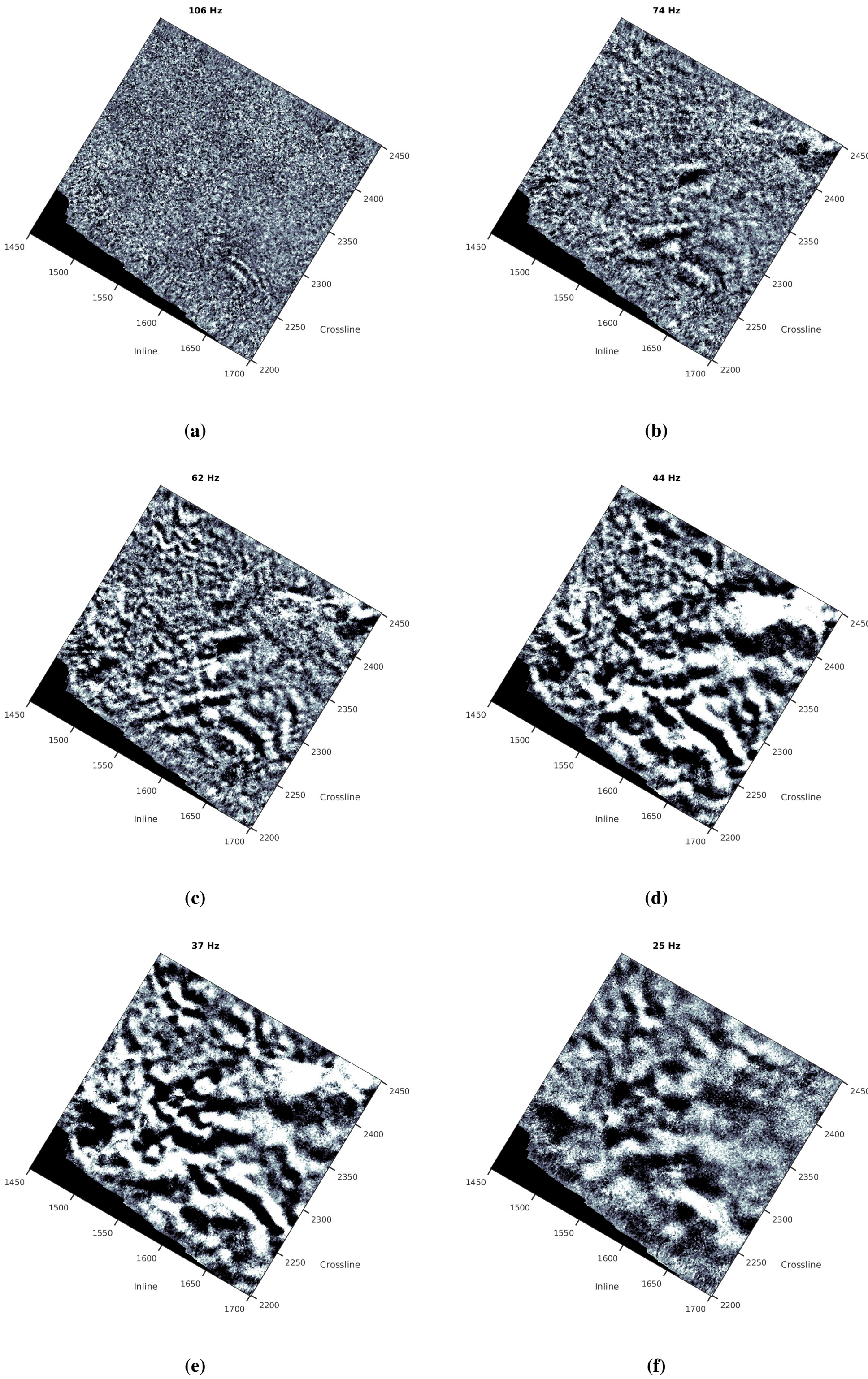


Figure 7: Spectral decomposed data for the geological horizon Top Stubensandstein. Iso-frequency slices range from 25 Hz to 106 Hz.

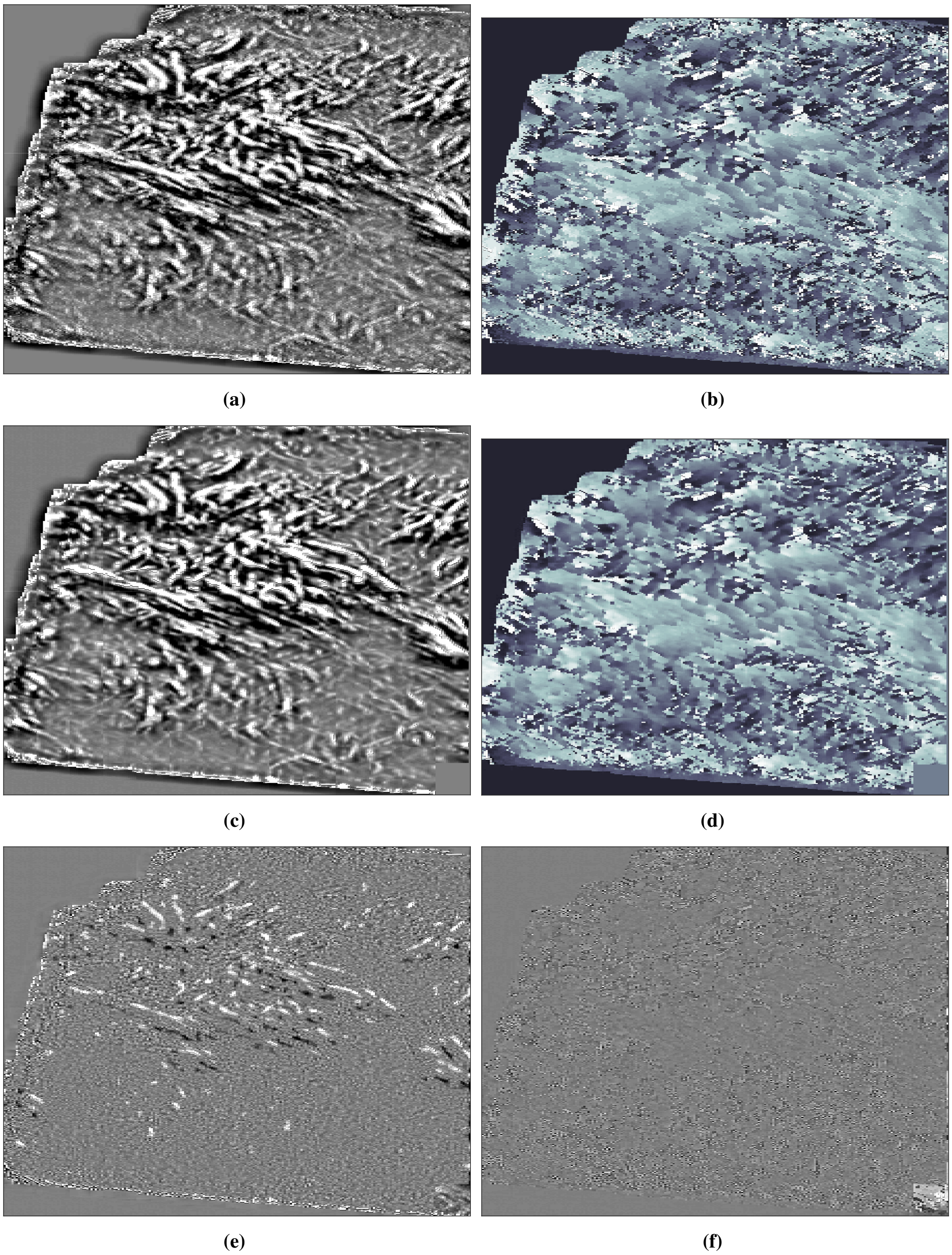
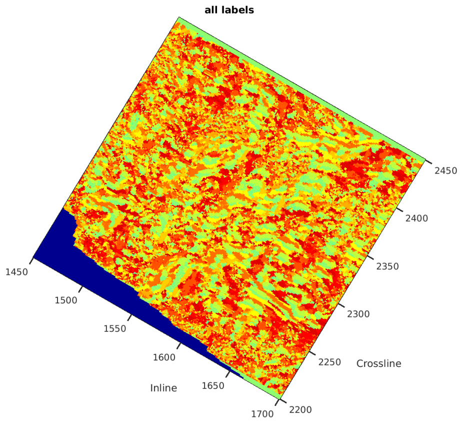
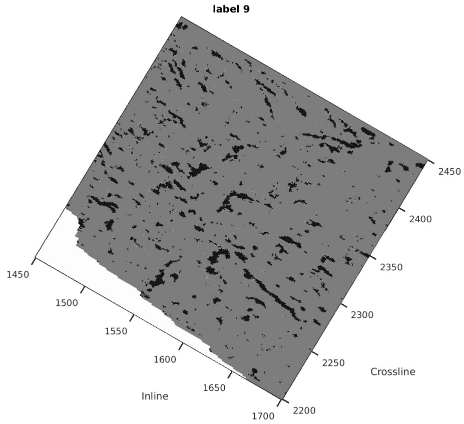


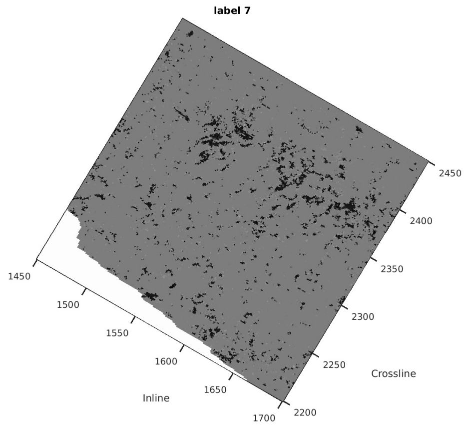
Figure 8: Result of encoding-decoding for diffractivity attributes for a time slice at 600 ms. Edge focusing is displayed in the left column: input image (a), reconstructed image (c), and the difference plot (e). Azimuth is displayed in the right column: input image (b), reconstructed image (d), and the difference plot(f). We observe a very good reconstruction of the input images. Note, the purpose of the autoencoder is not to perfectly reconstruct the image, i.e., learn an identity function, but to learn the underlying data generating representation. This representation should capture the most salient features observed in the reconstructed images, which explains the difference plots.



(a)



(b)



(c)

Figure 9: Labeled features for point and edge diffraction for the geological horizon Top Stubensandstein. a) All labels; b) Labels considered as edge diffractions indicating faults and fracturing; c) Labels considered as point diffractions indicating paleo-topography.

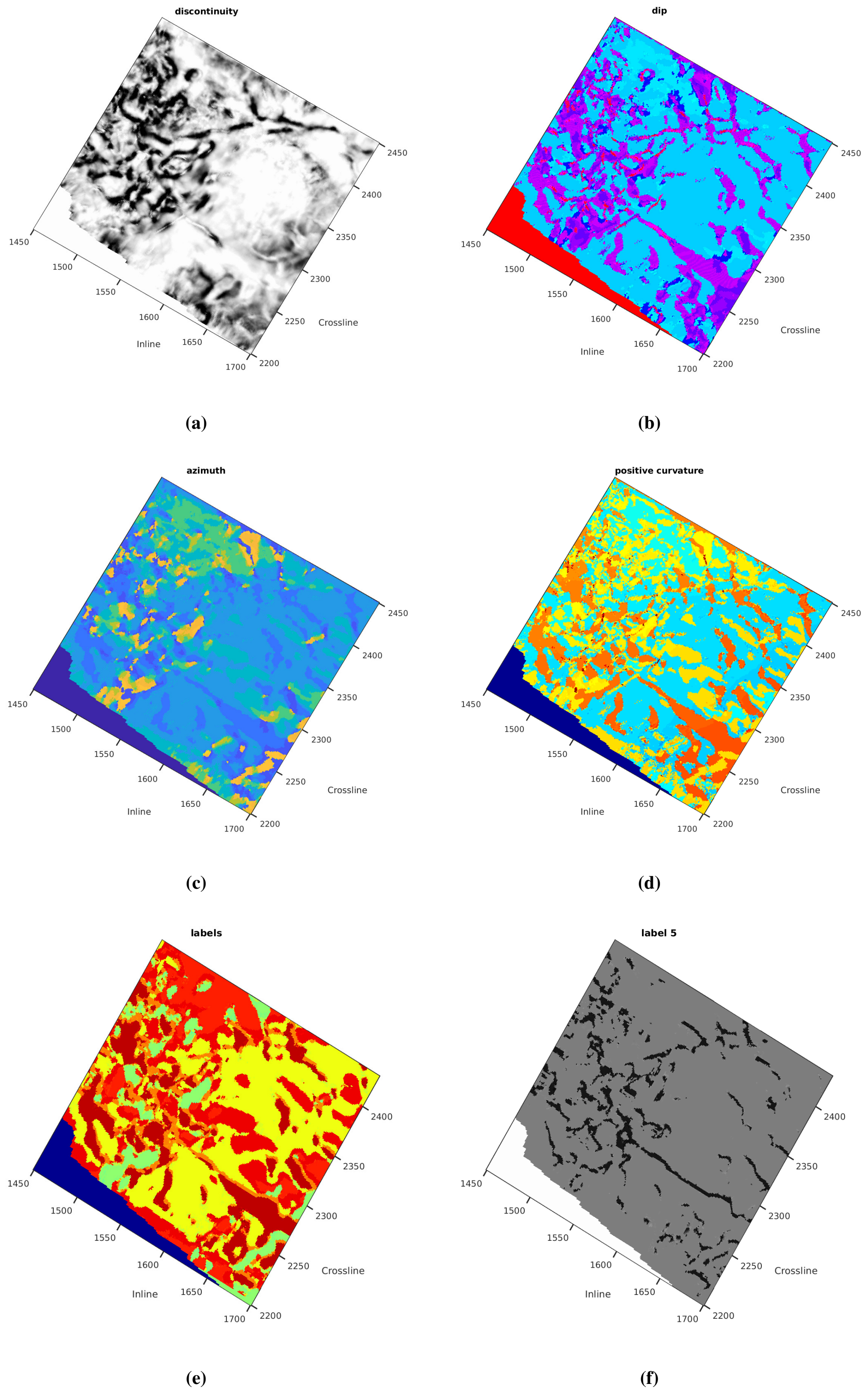
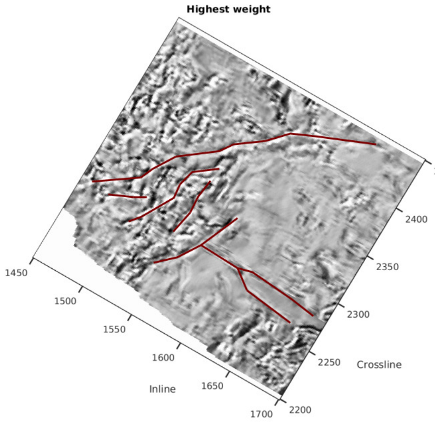
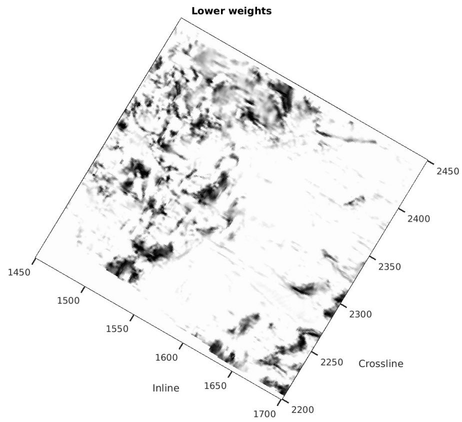


Figure 10: Geometric reflection attributes for the geological horizon Top Stubensandstein. a) discontinuity; b) dip; c) azimuth; d) the most-positive curvature; e) result of the unsupervised clustering (all labels) f) a labeled feature considered as potential edges.



(a)



(b)

Figure 11: Result of application of SNN to pre-trained networks. a) Sections with the highest correlation weights between diffractivity and reflection attributes. Comparing to Figures 9b) and 10f), the presumable discontinuities are more continuous as indicated by red lines. b) Sections with lower correlation weights between reflections and diffractivity attributes.