Neural Convolutional Seismic Amplitude Attribute Conv4_4-A: enhancing structural resolution

Abstract

Convolutional Neural Networks (ConvNet) have been created to process images and encode their features into parameters which opportunely trained can allow the ConvNet to recognize and classify maps similar to those which were presented to the network as ground true labels during the training phase.
Main application fields are in computer vision like face recognition, localization and detection for autonomous driving and robotics.
This process is called "supervised learning" because we teach the Network the kind of images it should recognize, classify, localize and detect.
In computer vision, training data are ideally bias-free. They are images or maps taken from a camera perhaps from different point of observation. In alternative they can additionally be extracted from a unique image through data-augmentation.

In 3D seismic, training and test data are difficult and expensive to collect.
Moreover, data still undergo a deterministic and linear processing and interpretation which can differentially "bias" the data, depending on the processing operator, on the choice of processing parameters and algorithms (survey geometry, trace editing, mute, deconvolution, spectral whitening, statics, v-analysis, wavelet/side lobes bandwidth filtering and tapering etc.).

The Geophysics operates within the "invisible environment" and is mainly involved in processing and interpretation of remote sensing signals.
Seismic processing, inversion, interpretation and attributes analysis are mainly based on linear models. Deep learning is gradually introducing the power of non-linearity also in these applications and this could open new perspectives to improve the accuracy of output results.
Another powerful potential for neural networks applications is the study of relationships among properties which are considered linearly independent. An example of this could be elastic and resistivity parameters like for instance rigidity and resistivity. Former studies developed deterministic methods to examine such relationships (A.Vesnaver, G. Mavko, A. Piasentin), neural networks now allows to derive local relationships among such parameters.

Recent developments in 3D seismic processing and inversion have been made to match models to the structural and rock-physics formations data of the subsurface.
Latest examples are Full Wave Inversion (FWI) and Reverse Time Migration(RTM), but these algorithms, like also prestack depth migration on the common image gathers, are still based on linear multivariate regression gradient descent, due to the actual limited computer capabilities.
Still here remain the indetermination of the intrinsic non-linearity and uncertainty present in every physical process, also considering that some methods are very sensitive to the initial input data of the velocity model.
Geophysicists will have to increase their effort to "un-bias" the processing and inversion algorithms and to develop more unsupervised classification algorithms and hybrid models to supply lack of training data.
Geophysics is a domain where the effort on algorithms development, due to the expensive measurements, has mostly been a priority.
The impact of neural networks will also bring additional powerful algorithms and will finally honor the intrinsic non-linearity within the problems to be solved.

Introduction

The method in this study can be denominated as supervised hybrid method because it takes advantage a powerful pre-trained ConvNet adapted to recognize a large variety of features but will use the trained parameters to elaborate the input maps combining an input of non-linear attributes to enhance resolution and indirectly parametrize the similarity of the input maps.

The results can also be analyzed through unsupervised methods, creating new features as a combination of the output data that can be useful to classify or improve the interpretation of the raw data which have been supplied as an input.

There are methods in computer vision that can have wide applications in geophysics.

The philosophy in Computer Vision, reveals similarities with the presumed visual human perception and visual knowledge process in which the image encoding originates from "singularities" perception.

In seismic attributes analysis the concept of singularities is related to the concept of waveform singularities and is defined by the Holder exponent **h** with values between -1 and ∞ which defines a Dirac Delta function (-1) a Heaviside unit step function (0) and a continuous analytical function (∞). (Li and Liner, Smythe, Chopra and Marfurt). One of the most common examples is instantaneous frequency from waveform interference (Taner et al.).
Further, the continuous function (h = ∞) can also be defined by local spatial 1^{st} and 2^{nd} derivatives.
Transferring the attribute from the functional space to the horizon map the concept can be extended into the visual perception defining singularities as deviations of the visual radiation frequency from a monotonic trend.

The parallelism between seismic attributes maps and images in the neural network convolutional process is also evident.
Seismic attributes are mostly represented in color-coded maps where color frequency is simply the numerical value of the attribute. How can we detect singularities with neural networks ?

Singularities parametrization in computer vision is performed through "Edge detection", where the Kernel parameters can be "learned" within the back propagation optimization process.
Therefore also in deep learning computer vision the dimension and dynamic of the receptive field detection drives the knowledge process.

When we talk about seismic interpretation, we talk about seismic inversion and the development of seismic attributes during and after the seismic processing phase.
These are parameters that are distributed within a 3D grid model, whose main intermediate task after seismic impedance will be to determine the elastic parameters of the formations. Their spatial configuration will determine the geological structural architecture of the reservoir under study.

Seismic attributes are derived from amplitudes, frequencies and phases components of the reflections at a certain reflector horizon. In seismic multicomponent they will consider also S waves components and their polarization.
The former classification subdivided attributes into: amplitude, complex and time attributes. The later development of research defined new concepts and sequences of parameters and each attribute was classified in function of the structural or rock-physical feature it would define (Marfurt, Chopra). Coherence, semblance and instantaneous frequency in thin layers mainly served to reconstruct the continuity and orientation of a reflector. Curvature, texture attributes are mainly used for anisotropy and fault detection. Spectral decomposition and wavelet transform are important for thin bed analysis to detect layers thickness and also elastic parameters, fluid type and mobility (Goloshubin).
In many applications, amplitudes attributes are translated into color frequency and intensity (gray-level), they were also used to predict texture matrix attributes (co-occurrence matrix): Energy, Entropy, Contrast, Homogeneity (Marfurt, Gao, West 2007) for further formational interpretations.
Complex attributes derived from amplitude real trace and its Hilbert transform (complex trace) are the main components of many interpretation methods.
Hampson-Russell implemented interpretation methods using complex attribute and derived parameters as linear applications in multiattribute linear regression analysis (Emerge).
A linear relationship was first calculated at the well location between seismic attributes and a target log to distribute in all the 3D seismic volume. The "less linear" attributes were discriminated through

a “cross-validation” procedure and a series of “regression coefficients” were derived for the log calculation through the 3D cube.

In this method mainly amplitude attributes were used, derived from the real trace, the complex trace in the imaginary plane and their envelope (signal strength) (Figure 1). The same parameters are used in neural networks supervised methods as features in training examples.
The main derived complex amplitude components used in these methods are: Signal strength, instantaneous phase, instantaneous frequency, average and dominant frequency, amplitude weighted cosine phase, amplitude weighted frequency, amplitude weighted phase, cosine instantaneous phase, apparent polarity, derivative instantaneous amplitude, second derivative instantaneous amplitude, integrated absolute amplitude.
The successfully predicted target logs were: P-wave velocity, porosity, density, gamma ray, water saturation, lithology (seismic impedance).
It was anyway the introduction of neural networks that fully honored the non-linearity of the prediction process with training data (attributes) and labels (target logs) in supervised applications. At the beginning fully connected neural networks were used and these were also successful in predicting missing logs and petrophysical properties on the 3D cube with improved accuracy compared to linear multivariate regression methods.

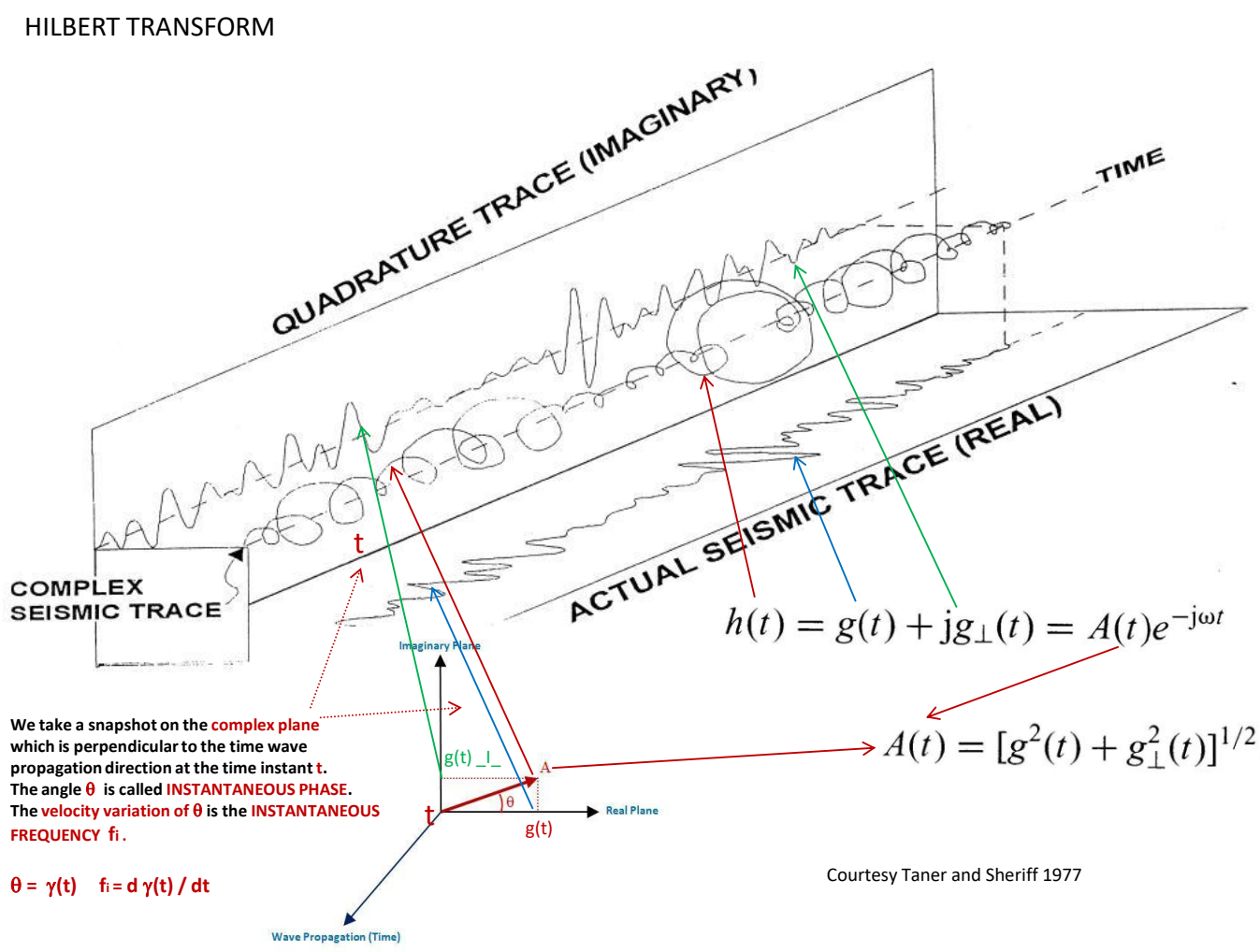


Figure 1 *Amplitude complex attributes (Taner et al., 1979) (readapted A. Piasentin 2016)*

Method and Theory

Color codes are widely used to visualize interpretation parameters (structural attributes, spectral decomposition, AVO rock physics attributes and further wide range of attributes like the above mentioned amplitude attributes). Color codes use three classes in RGB pixel intensity values to better visualize the attribute quantification for the interpretive phase.
The best example in this context is the visual presentation of seismic impedance sections after prestack and poststack seismic inversion.
For visual interpretive scope, parameters like seismic impedance, $\lambda\text{-}\rho$ $\mu\text{-}\rho$ **, Vp/Vs** or partial stack amplitudes in AVO analysis are mainly translated into colors and mapped.
In Computer Vision colors are also coded as RGB intensity values which range between 1 and 255 for each base color.

In computer vision applications we can build a Rank-3 Tensor of the 3 base colors one for each channel representing increasing color frequencies.
Each base color can be mapped into one singular channel and the superposition of the three channels can allow further processing during the training phase.
Within these building blocks of 3x255 pixel frequencies values, we can build every kind of high resolution images reproducing natural colors with the highest resolution.

The workflow consists of preprocessing and normalize the seismic data after seismic processing and inversion to values between 1 and 255 for each RGB "radiation intensity" to visually reproduce the quantification of the specific attribute on the seismic map. The program will sequentially extract normalized amplitudes. These can be further processed into a continuous frequency and amplitude map.
The input of two differentiated amplitude maps into the network provides the resolution enhancement due to the activation differentiation within the neural network.
By examining the activations at channels outputs it was possible to observe the enhanced resolution.
Input attributes can also be preferentially non-linear attributes and this can result by using bin and superbin reflections or randomly shuffling the fold and subdividing reflections into 2 attributes classes.

Visual Interpretation and Parametrization

Visual representation of seismic attributes offers the possibility to an expert interpreter for direct interpretation of the data. However parametrization of the data within the neural networks processing offers the possibility to calculate additional parameters that give further indications on the structural and rock-physical properties of the formations and diagnostic attributes during the seismic processing and inversion process. This can be important for a quality control on the fold reconstruction, to examine which reflections belong to the same bin, for v-analysis and migration. Other critical applications regard the parametrization of anisotropy and texture attributes like entropy, energy, contrast. Finally, parametrization can be integrated in the seismic imaging process, prestack depth migration, RTM, FWI, prestack and poststack seismic inversion.
In this context, convolutional neural networks applications have the potential to improve interpretation methods, because they are supported by a wide range of available optimization algorithms with the flexibility of processing workflows and the parallelism of operations offered by the available programming frameworks.

For this study “Tensorflow” and partially “Keras” was used for program implementation.
These are high-level API (programming frameworks), written in Python and capable of running on top of several lower-level frameworks. Keras itself, as API running on top of Tensorflow, represents one of the highest level frameworks with eager execution.

Application of Convolutional Neural Network for the Seismic Attributes Mapping

This study presents a method to combine two or more different attributes to enhance resolution and calculate further elastic properties to improve 3D seismic interpretation. Sequentially two tensors (STRUCT1 and STRUCT2) of different amplitude ranges are input into the VGG19 model.
The two different amplitude attributes belong to coincident bin numbers of a reflector slice.
The goal was to supply these attributes as input into the ConvNet VGG19 convolutional neural network to enhance the resolution and edge positioning of the structures identified after the seismic inversion.
This first study presents the visual superposition of the effects with enhanced resolution.
This illustrates a methodology that can be extended to other sorts of attributes for quality control.
The name used for this output parameters is " Neural Convolutional Seismic Amplitude Attribute Conv4_4-A" to emphasize the "A" algorithm that creates them and the processing level at the stage of layer Conv4_4 output of a VGG19 architecture where it is produced.
The network starts by analyzing and encoding on the first layers, the low level features at the single bin resolution.
The attribute map undergoes through convolutional and non-linearity RELU operations at each layer.
The Network consists of a sequence of Convolutional and Max pooling layers.
The filters are all of dimension f=3x3 and stride=1 and "same convolution" is used.
All Max pooling have a width of 2x2 and a stride of s=2. This reduce the dimensions in output of the next layer by a factor 2. Therefore each further layer after pooling will have half the n_h, n_w dimensions of the previous layer.
The dimensions of layers activations here reported are the original ones reported in the original model (Symonian & Zisserman 2015) and represent the proportional decrease at each layer block due to the Max-pooling effect.

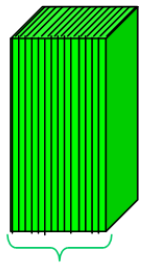
Original VGG19 Model description

input_1 map 224x224→ 64(3x3)filters - same convolution → 224x224x64
block1_conv1 224x224 - 64 channels
block1_conv2 224x224 - 64 channels
block1_pool f=2 s=2
block2_conv1 112x112 - 128 channels
block2_conv2 112x112 - 128 channels
block2_pool f=2 s=2
block3_conv1 56x56 - 256 channels
block3_conv2 56x56 - 256 channels
block3_conv3 56x56 - 256 channels
block3_conv4 56x56 - 256 channels
block3_pool f=2 s=2
block4_conv1 28x28 - 512 channels
block4_conv2 28x28 - 512 channels
block4_conv3 28x28 - 512 channels
block4_conv4 28x28 - 512 channels
block4_pool f=2 s=2
block5_conv1 14x14 - 512 channels
block5_conv2 14x14 - 512 channels
block5_conv3 14x14 - 512 channels
block5_conv4 14x14 - 512 channels
block5_pool f=2 s=2

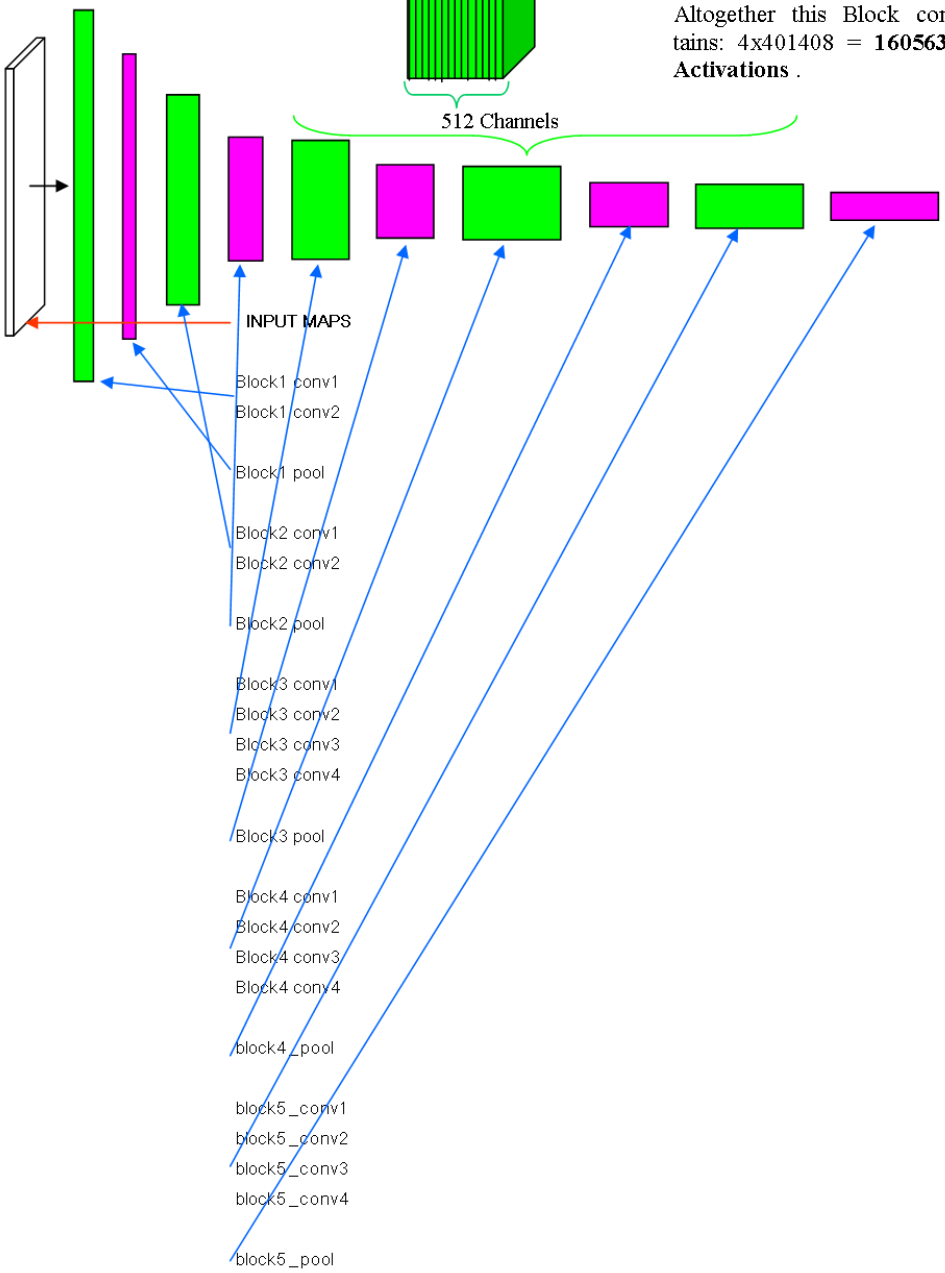
Pooling Layer used are Max pooling with f=2 and Stride=2 which will reduce the dimension of a factor 2. Non-linearity function used is RELU, which provides flexibility and speed to the program execution. After each filter operation, a bias is summed (broadcasted) to the resulting channel and finally the RELU non-linearity is applied to form each channel of the next layer.

VGG19 ConvNet Architecture :
Schematic of the original model

Block 4 _ conv 4
512 channels
Each channel 28x28



Volume: Block 4 conv 4
Higher level features are analysed in this layer to focus on enhanced resolution .
Block 4 is composed of 4 ConvNets layers each containing **512 channels** of dimensions **28x28.**
Altogether this Block contains: 4x401408 = **1605632 Activations** .

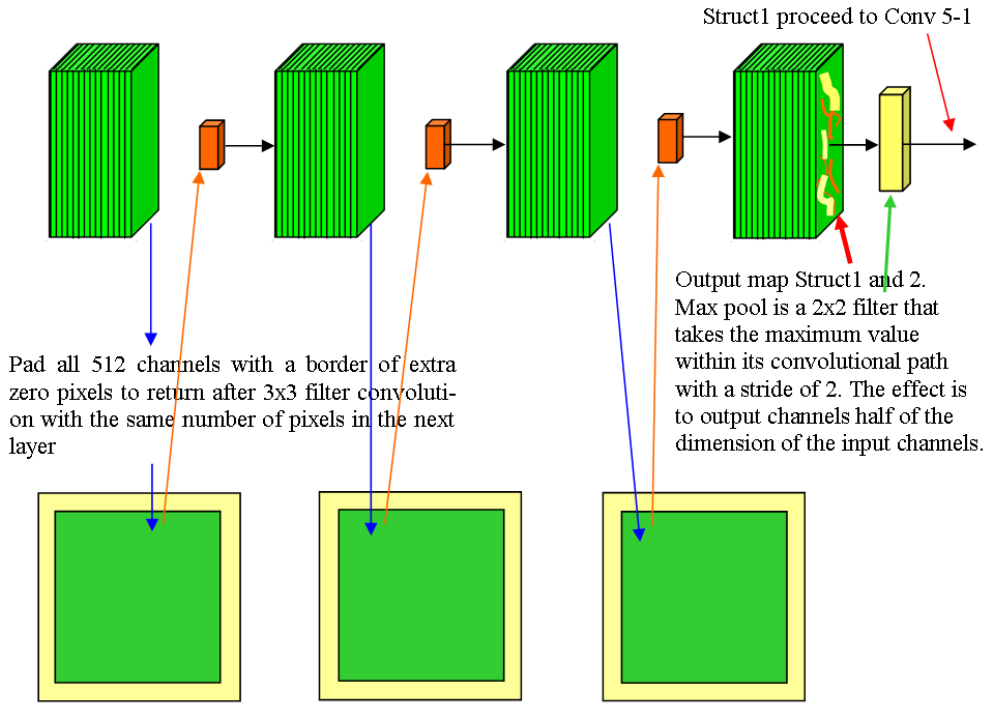


A. Piasentin

SCANNING BLOCK 4

Block4-Layer4 will output in this program the covariance attributes for structure refinement. Whereas in computer vision inputs and outputs of algorithms are pixel maps, 3D seismic analysis will input and output the network with maps of attributes per bin, or reflection specific attributes after specific processing, inversion and attributes calculation.

In the original model Block 4 contains 4 layers, each one with 512 channels. The output of this layer is used as structure resolution refinement.
Each layer undergoes „same" convolution. Therefore each channel is padded with external zeros and undergoes subsequent filter convolutionals operations with 512 filter each of dimension 3x3 plus bias and RELU non-linearity finally applied.
The next layer will have therefore the same number of bins and the same number of channels. All 4 layers in this block undergoing these operations will have the same dimensions.



A. Piasentin

Activation function RELU

The non linearity function is a RELU (Rectified Linear Unit) with activation a = max(0,Z).
The derivative is 1 for Z >= 0 and 0 for Z < 0 .
This activation with simple derivatives calculations will speed up gradient descent (Figure2).

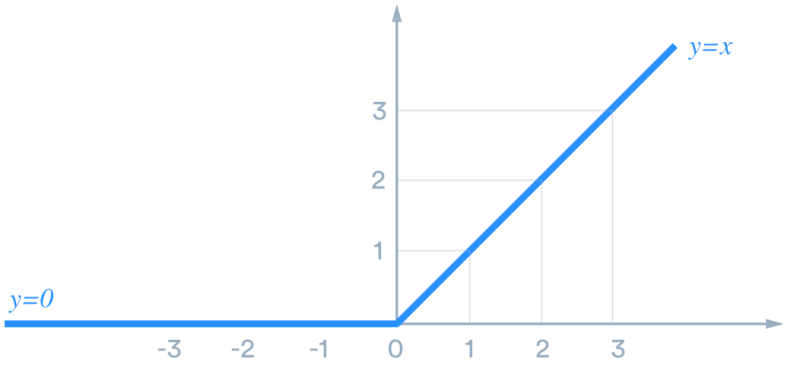


Figure 2 *RELU function adds non-linearity to produce the layer activation, used in ConvNet substituting the previously used Tanh function*

Examples of Differentiated Amplitude maps

This is an example of a color visualization of differentiated amplitude maps to input additional structural information to the ConvNet (Figure 3 and Figure 4).

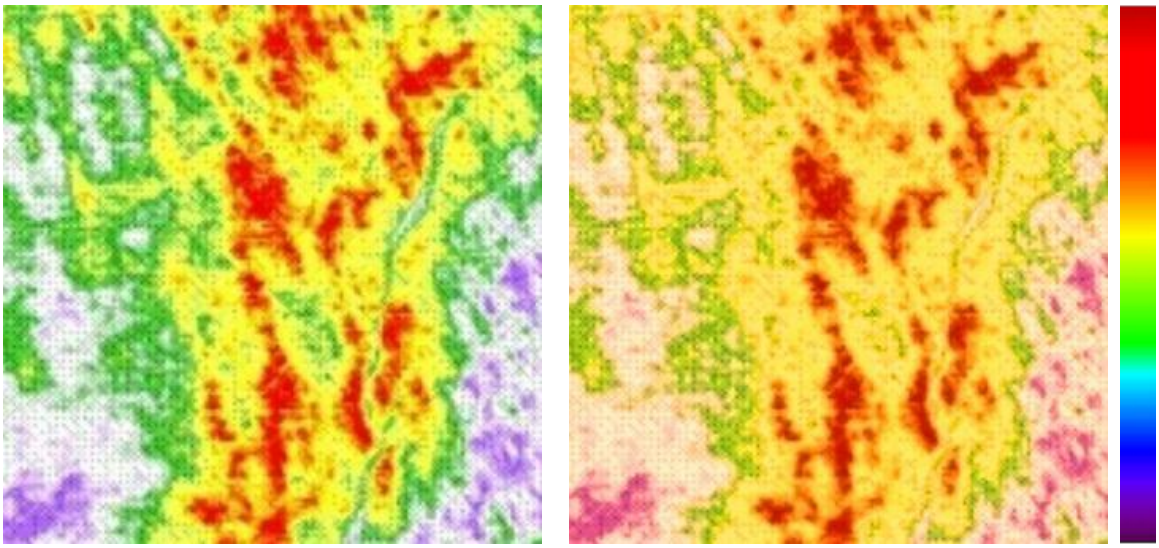


Figure 3 and 4 *Two differentiated amplitude input maps STRUCT1and STRUCT2. Highest amplitudes are presented in red , lowest amplitudes in violet.*

The amplitude map STRUCT1 was used as input to the VGG19 ConvNet together with the scaled amplitude map STRUCT2 to produce the output high resolution amplitude attribute .

Output map

The generated map (output map) is initialized with random amplitudes.
Sequentially two digital amplitude maps of common bin reflections, belonging to the same reflector horizon but each with individual amplitude content, are input into the ConvNet. Each of them undergo different processing steps.
The structural affinity is reached minimizing the output layer activations of the input image and the generated image. The higher resolution is reached by minimizing the unnormalized cross-covariance between activations between adjacent channel of the second input map and the generated map.
(Figure 5).

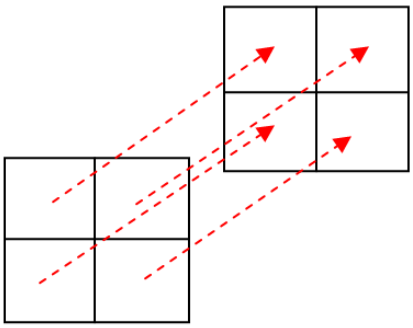


Figure 5 *Schematic cross-covariance between activations across channels*

Loss Function and Cost Function

STRUCT1 loss function
This is the sum of the difference of activations in corresponding layers between the generated map initially implemented as white noise and of the input map STRUCT1. During gradient descent this difference is minimized.

$$J_{\_struct1} = \frac{\sum_{activations} \left(a^{(P)} - a^{(S)} \right)^2}{d}$$

Equation 1

$a^{(P)}$ = activation of layer Conv4_4 in the map STRUCT1
$a^{(S)}$ = activation of corresponding layer in the output generated map
d = normalization factor for the total number of (activations x channels)^2 * 4

STRUCT2 Loss Function

This is calculated from the unnormalized cross-covariance matrix across activations of nearby channels in corresponding layers of the input map of STRUCT2 and the output generated map.
covariance matrix is of n_c x n_c dimension (n_c = number of channels in the layer where the matrix

is calculated). For this work the algorithm was applied to layer Conv 4_4 of STRUCT2 and of the output map.

During gradient descent the difference in covariance matrices $G_{hk}^{(P)}$ and $G_{hk}^{(S)}$ are minimized, producing the output attribute on the generated map with enhanced structural resolution (Equation 2).

$$J\_struct2 = \frac{\sum_{h=1}^{c}\sum_{k=1}^{c}\left(G_{hk}^{(P)} - G_{hk}^{(S)}\right)^2}{d}$$

Equation 2

$G_{hk}^{(P)}$ = unnormalized cross-covariance matrix for layer conv 4_4 of struct2 map

$G_{hk}^{(S)}$ = unnormalized cross-covariance matrix for layer conv 4_4 of output generated map h,k = height and width of channels

c = number of channels

d = normalization factor for the total number of (activations x channels)^2 * 4

Total Cost Function

The hyperparemeters α and β are used to enhance the contribute of input attribute STRUCT1 or STRUCT2 on the output generated map.

During gradient descent the total cost function is minimized in order to bring together the effect of the two input attributes in order to generate the new covariance attribute in the output map.

$$J\_Cost = \alpha * J\_struct1 + \beta * J\_struct2$$

Equation 3

J_Cost = Total loss / cost function

J_Struct1 = loss function of STRUCT1

J_Struct2 = loss function of STRUCT2

α = Hyperparameter , weight of the loss function of STRUCT1

β = Hyperparameter , weight of the loss function of STRUCT2

Program Optimization

For the model implementation Adam optimization was used with 500 Epochs on minibatches. Adam is a very flexible optimization algorithm which speed up the calculations and showed a great flexibility on a various range of neural networks architectures. It combines gradient descent with Momentum and RMSprop, by taking advantage of running averages on the gradients, and optimizing the directivity versus the global minimum of the cost function. It provides smoothing the orthogonal directivity components on gradient descent iteration steps, smoothing oscillations and concentrating it to the maximum gradient direction. This also allows to use larger learning rates.

α and β in this new context (in italic type) are different hyperparameters as those used before in the total cost function .

The first step is the implementation of the momentum computation is expressed by the following equations 4:

$$v_{dW^{[l]}} = \beta v_{dW^{[l]}} + (1-\beta)dW^{[l]}$$
$$W^{[l]} = W^{[l]} - \alpha v_{dW^{[l]}}$$
$$v_{db^{[l]}} = \beta v_{db^{[l]}} + (1-\beta)db^{[l]}$$
$$b^{[l]} = b^{[l]} - \alpha v_{db^{[l]}}$$

Equations 4

l varies from 1 to L where L : the number of layers

β : hyperparameters that control the exponentially weighted averages

α : learning rate

dW : derivative of cost function with respect to w

db : derivative of cost function with respect to b

Adam optimization is a combination of Momentum and RMSprop.

First exponentially weighted average of past gradients are computed and stored in a variable $\boldsymbol{v}$ (momentum), then bias correction is performed to get: $\boldsymbol{v^{corrected}}$

Bias correction is applied to avoid a defect of the running averages algorithmus where the initial part of the function is "biased" in excess with respect to the real average values.

In a second step, the exponentially weighted average of the squares of the past gradients are computed and stored in a variables $\boldsymbol{S}$ (RMSprop).

After performing bias correction in $\boldsymbol{S}$ we can get: $\boldsymbol{S^{corrected}}$

Momentum →
$$v_{dW^{[l]}} = \beta_1 v_{dW^{[l]}} + (1-\beta_1)\frac{\partial \mathcal{J}}{\partial W^{[l]}}$$
$$v^{corrected}_{dW^{[l]}} = \frac{v_{dW^{[l]}}}{1-(\beta_1)^t}$$ ← **Bias correction**

RMSprop →
$$s_{dW^{[l]}} = \beta_2 s_{dW^{[l]}} + (1-\beta_2)\left(\frac{\partial \mathcal{J}}{\partial W^{[l]}}\right)^2$$
$$s^{corrected}_{dW^{[l]}} = \frac{s_{dW^{[l]}}}{1-(\beta_1)^t}$$ ← **Bias correction**

Adam →
$$W^{[l]} = W^{[l]} - \alpha \frac{v^{corrected}_{dW^{[l]}}}{\sqrt{s^{corrected}_{dW^{[l]}}} + \varepsilon}$$

Equation 5

The update rule is, for $l=1,...,L$

L: number of layers

β_1 and β_2: hyperparameters that control the two exponentially weighted averages

α: learning rate

ε: is a very small number to avoid dividing by zero

Output results

The result show an improved resolution of the amplitude attribute in the studied area defining additional structural lineaments (Figure 6).

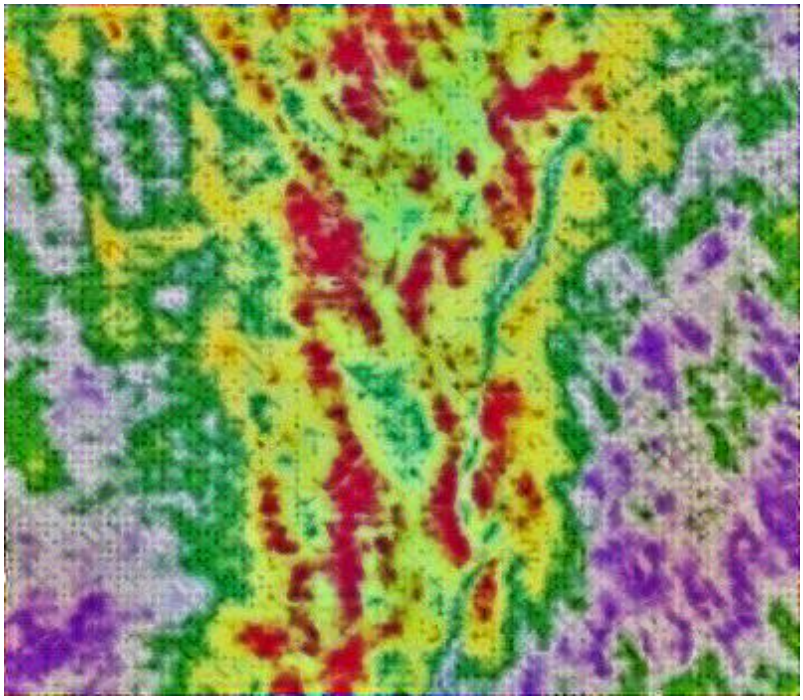


Figure 6 . *This is the output of the ConvNet optimization processing, with evident improvement in the 3D seismic attribute resolution*

Application of structural filtering to the output map

Structural filter: black/white high resolution (Figure 7 and Figure 8).
The black-white map shows the output map with enhanced structured filtering in the main anisotropy direction and compares the input image before the convolutional neural network processing with the image in output of the ConvNet used for this study.

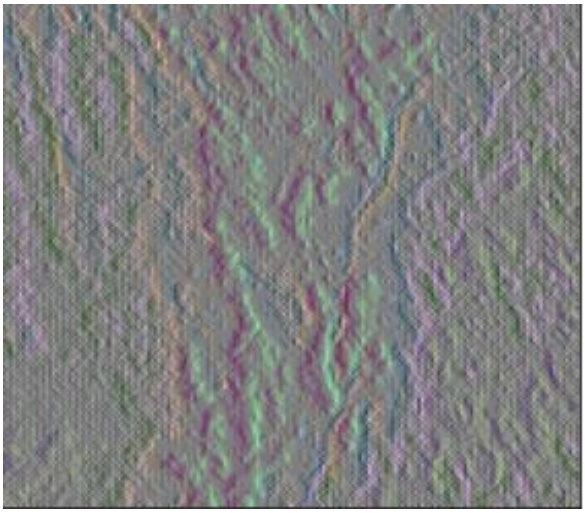


Figure 7 *Input image before ConvNet processing*

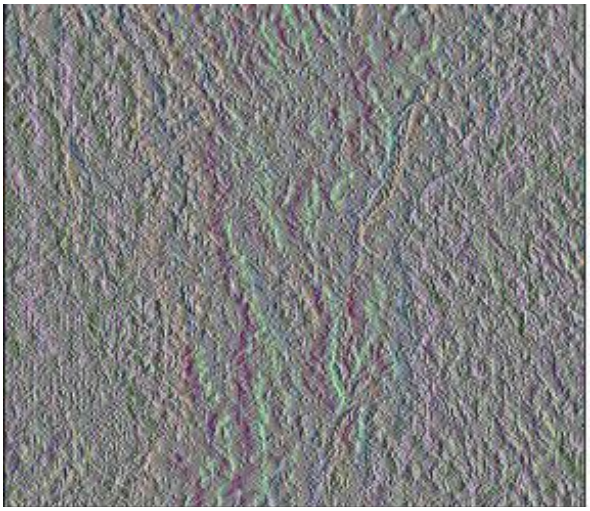


Figure 8 *Output image after ConvNet processing and filtering with enhanced resolution*

The resolution have increased. This higher frequency component represents an edge-detector.

Conclusions

This study is a provisional result in one research field conducted at the Neural-Geophysics Lab in Augsburg.
Based on different channel operations on features identification, structural differentiation is performed and transmitted in output to the generated map, enhancing this way the resolution of the amplitude attribute.
This shows how combining 3D seismic attributes through a convolutional neural networks, a new attribute with improved resolution can be created.
New studies with improved algorithms shows even more powerful capabilities by combining seismic and petrophysical attributes with the goal of improving resolution in mapping of subsurface structures and their inherent rock physical and petrophysical parameters. These studies can have an important impact in seismic/petrophysical reservoir characterization and quality control.

Acknowlegments

The concept, transfer learning and helper codes for this study were inspired after modification from previous versions of the programs and papers developed and written by: A. Krizhevsky, I. Sutskerev, G. Hinton, K. Simonyan, A. Zisserman, L.A. Gatys, A. S. Ecker, M. Bethge.
Later contributions from Billy Lamberta, Mark Daoust, Minho Heo, Yash Kataryia were also considered.
Mostly helpful were all theoretical concepts for theory and programs implementations in “deep learning” developed by: A. Ng, K. Katanforoosh, Y. Bensouda Mourri, A. Kiani, B. Uyumazturk of the Coursera Community.
Papers and programs in the TensorFlow Hub and Google Developer Site were reviewed and were useful to develop the new concept for this study.
Of fundamental help were all research papers of the CREWES research consortium in Calgary and papers and tutorials from: R. Garotta, K. Marfurt, S. Chopra, G. Ballay, D. Hampson, P. Van der Smagt.

References

Taner, M.T., Koehler, F., and Sheriff, R.E. [1979].
Complex Seismic Trace Analysis:
Geophysics , 44 (6): 1041-1063.

Taner, M.T. and Sheriff, R.E. [1977].
Application of Amplitude, Frequency, and Other Attributes to Stratigraphic and Hydrocarbon Determination:
In American Association of Petroleum Geologists Memoir, Vol. 26, 301.

Satinder Chopra and Kurt J. Marfurt. [2007].
Seismic Attributes for Prospect Identification and Reservoir Characterization:
SEG : Society of Exploration Geophysicists.

Dan Hampson, Jim Schuelke, John Quirein. [2000]
Using multi-attribute transforms to predict log properties from seismic data:
SEG : Society of Exploration Geophysicists.

Brian H. Russell. [2005].
Introduction to Seismic Inversion Methods:
SEG : Society of Exploration Geophysicists.

Margrave, G.F., Stewart, R.R., Larsen, J-A. [2001].
Joint P-P and P-S seismic inversion:
CREWES Research Report — Volume 13 (2001).

Larsen, J.A. [1999].
AVO Inversion by SimultaneousP-P and P-S Inversion:
M.Sc. Thesis University of Calgary Department of Geology and Geophysics, September, 1999.

Mavko, G., Mukerji, T., and Dvorkin, J. [1998].
The Electrical Resistivity - Acoustic Velocity Relationship: A Method for Constraining Porosity:
The Rock Physics Handbook, Springer.

Vesnaver A. [1994].
Towards the uniqueness of tomographic inversion solutions:
Journal of Seismic Exploration · January 1994

Piasentin A. [2014].
Deterministische und stochastische Methode zur Berechnung, Verbreitung und Interpretation petrophysikalischer Mikro-Eigenschaften und Makro-Migenschaftenals petrophysikalisch-seismische Attribute im „3D Seismic Volume“ für Interpretationszwecke:
http://angelo-piasentin.eu/patent-de-102014000234a1-offenlegung.
Offenlegungsschrift DE102014000234A1 2015.07.09
Deutsches Patent- und Markenamt

Krizhevsky, A., Sutskever, I., Hinton, G. [2012].
Imagenet classification with deep convolutional neural networks:
In: NIPS (2012)

Karen Simonyan & Andrew Zisserman [2015].
Very deep convolutional networks for large-scale image recognition:
Published as a conference paper at ICLR 2015

Leon A. Gatys, Alexander S. Ecker, Matthias Bethge [2015].
A Neural Algorithm of Artistic Style:
arXiv.org > cs > arXiv:1508.06576 .

Patrick Van der Smagt [2017].
Machine learning and Neural Networks – Course Tutorials:
GeoNeurale 2017.

Frueh, H., Perucchi, T. [2007]
Grundkurs Neuronale Netze – Course Tutorials:
Neuronics - Zurich 2007.

Simonyan, K., Vedaldi, A., Zisserman, A. [2013]
Deep inside convolutional networks: Visualising image classification models and saliency maps:
arXiv 1312.6034v1 (2013).

Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. [1989].
"Backpropagation applied to handwritten zip code recognition:
" *Neural Computation*, vol. 1, no. 4, pp. 541–551, Winter 1989.

Y. LeCun, L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, and W. Hubbard [1989].
"Handwritten digit recognition: Applications of neural net chips and automatic learning:":
IEEE Trans. Commun., vol. 37, pp. 41–46, Nov. 1989.

O. Matan, H. S. Baird, J. Bromley, C. J. C. Burges, J. S. Denker, L. D. Jackel, Y. LeCun, E. P. D. Pednault, W. Satterfield, C. E. Stenard, and T. J. Thompson. [1992].
"Reading handwritten digits: A ZIP code recognition system:
" *IEEE Trans. Comput.*, vol. 25, no. 7, pp. 59–63, July 1992.

Y. Bengio and Y. LeCun. [1994].
"Word normalization for on-line handwritten word recognition:
" in *Proc. IEEE Int. Conf. Pattern Recognition*, Jerusalem, 1994.

[93] R. Vaillant, C. Monrocq, and Y. LeCun. [1994].
"Original approach for the localization of objects in images:
" *Proc. Inst. Elect. Eng.*, vol. 141, no. 4, pp. 245–250, Aug. 1994.

D.C. Cire¸san, U. Meier, J. Masci, L.M. Gambardella, and J. Schmidhuber. [2011].
High-performance neural networks for visual object classification:
Arxiv preprint arXiv:1102.0183, 2011.

A. Krizhevsky. [2009].
Learning multiple layers of features from tiny images:
Master's thesis, Department of Computer Science, University of Toronto, 2009.

Y. LeCun, K. Kavukcuoglu, and C. Farabet. [2010].
Convolutional networks and applications in vision:
In Circuits and Systems (ISCAS),
Proceedings of 2010 IEEE International Symposium on, pages 253–256. IEEE, 2010.

C. Szegedy, A. Toshev, and D. Erhan. [2013].
Deep neural networks for object detection:
In C. J. C. Burges, L. Bottou, Z. Ghahramani, and K. Q. Weinberger, editors,
NIPS, pages 2553–2561, 2013.

Le, Q. V., Ngiam, J., Chen, Z., Chia, D., Koh, P.W., Ng, A.Y. [2010].
Tiled convolutional neural networks:
Computer Science Department, Stanford University.

Zeiler, M., Taylor, G., Fergus, R. [2011].
Adaptive deconvolutional networks for mid and high level feature learning:
In: ICCV (2011).