Combining classical processing with Machine Learning

CREWES 2022

Daniel Trad

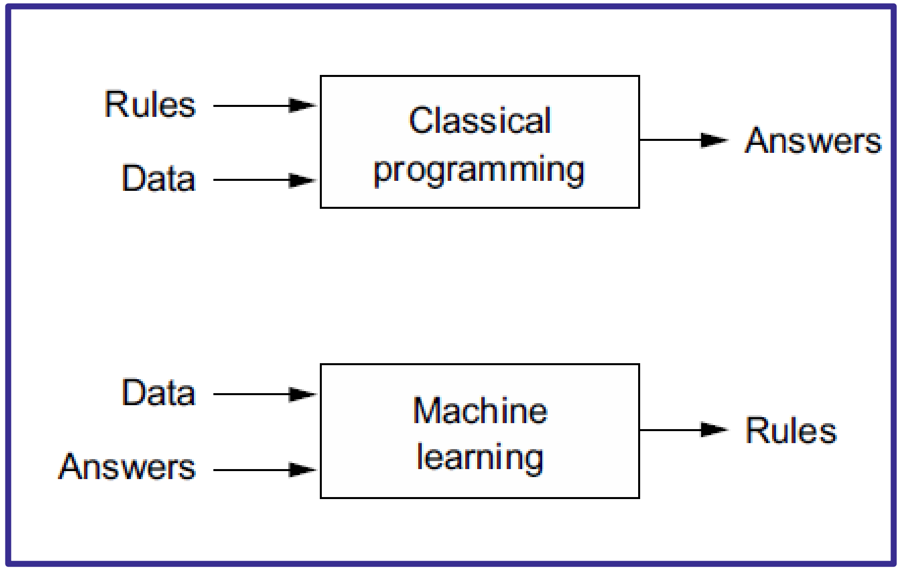
December 2, 2022

NSERC CRSNG
UNIVERSITY OF CALGARY
FACULTY OF SCIENCE
Department of Geoscience

Outline

- Physics vs Machine Learning
- Combining Seismic Processing and Deep Learning
  - Least Squares Migration
  - Radon multiple attenuation
  - Ground Roll attenuation
  - Interpolation
- Conclusions

Physics and Machine Learning different approaches



Francois Collet, 2021
Machine Learning: a new programming paradigm

Both **Physics** and **Machine Learning** learn from experiments and observations
Physics needs **rules** and **ML** doesn't.

Every experiment has a many possible outcomes.

Physics uses rules to select outcomes that matter.
This is called **sparsity.**

ML lets anything to be learned but **prunes** by **training**

AI requires more **computer power** and **more data** to do pruning but it is more **flexible**.

Daniel Roberts, 2021, why is AI hard and physics simple?

In recent years, a combination of the 2 approaches called Physics Guided ML is taken momentum.

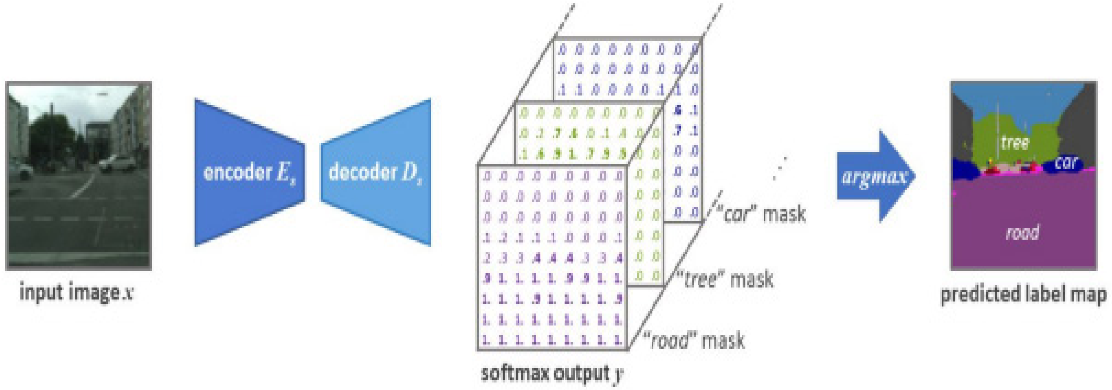
Computer Vision Applications

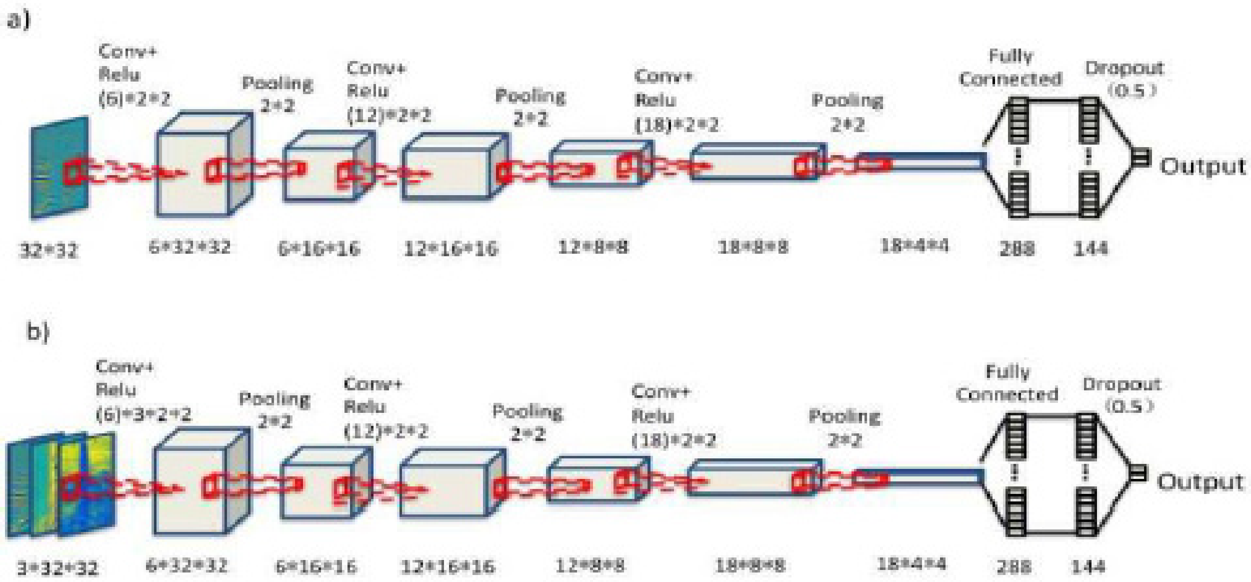
Denoising

Super resolution

Semantic segmentation

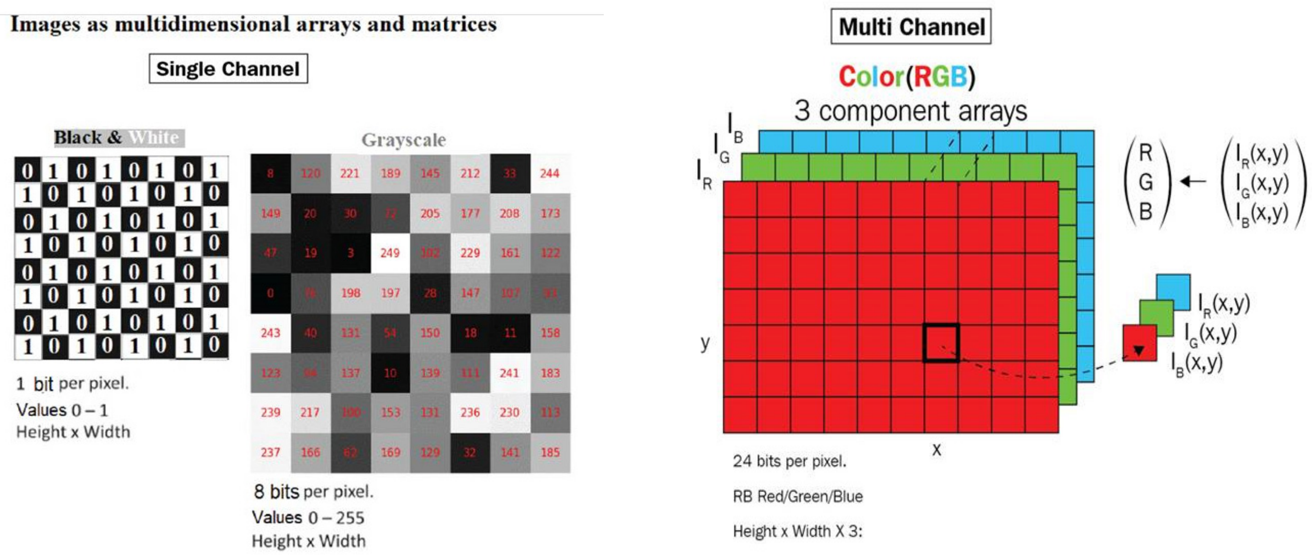
Géron, Aurélien. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow . O'Reilly Media.





Schuster, Practical ML in Geosciences, 2021

Computer Vision vs Seismic Processing

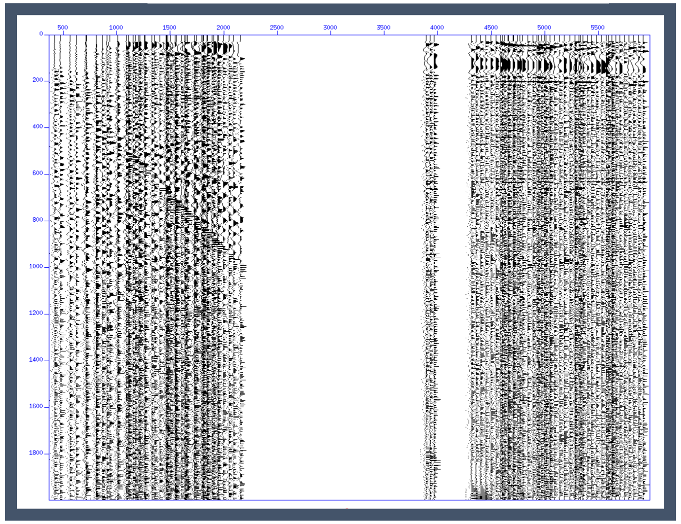


The world according to computer vision

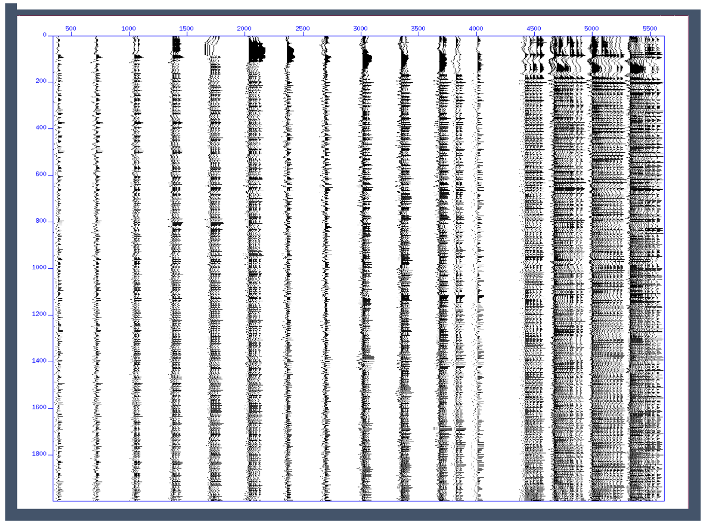
Sandipan Dey. Hands-On Image Processing with Python. Packt.

The world according to seismic processing: the same 3D shot record…

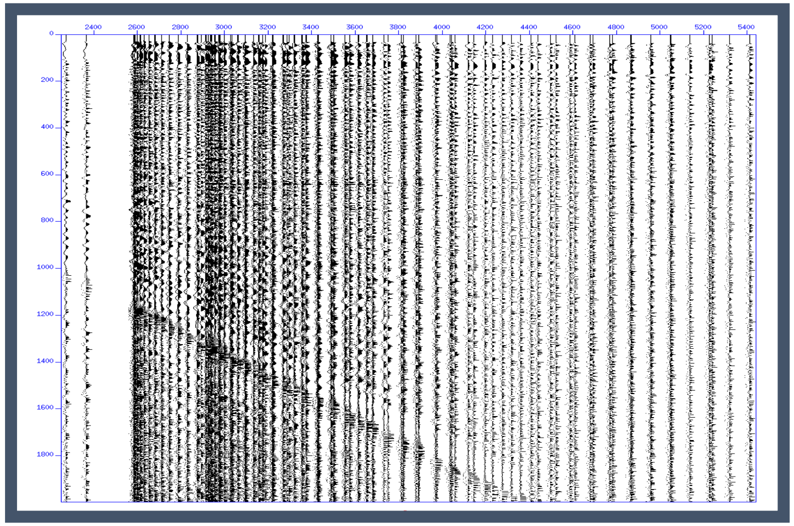
sorted by offset



just a few meters appart



organized by lines

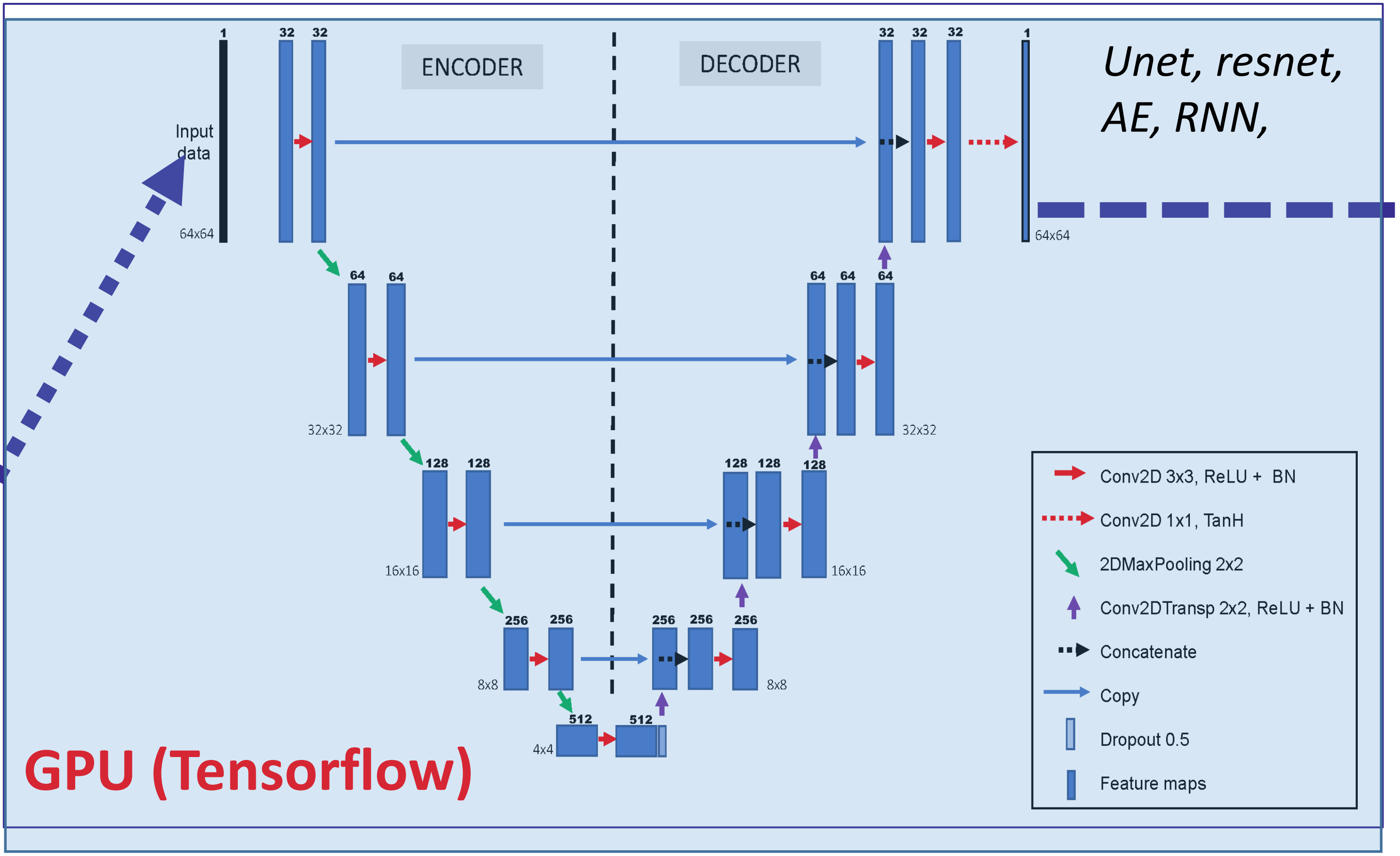


Combination of standard processing with DL

GPU (CUDA)

Madagascar, multi-environment processing flow with "scons"

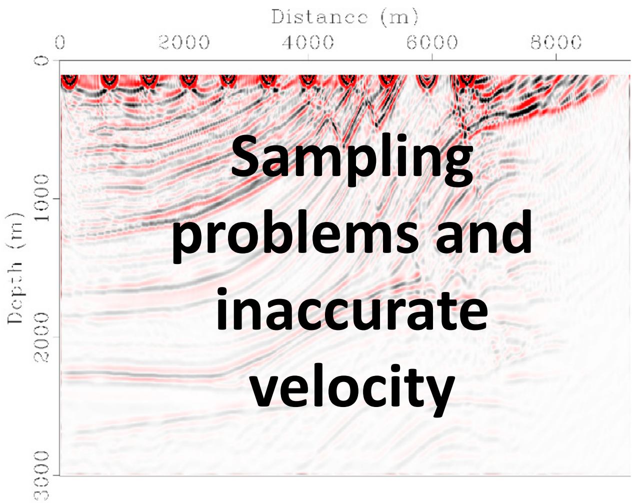
- modeling
- processing
- migration or transforms
- machine learning
- inverse transforms
- processing
- Interpretation



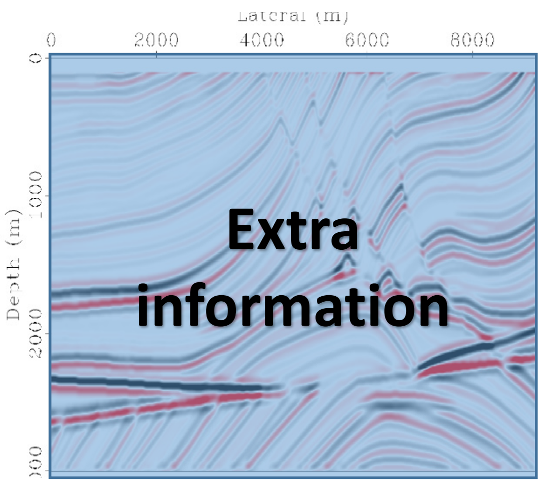

```
Total params: 3,839,841
Trainable params: 3,835,937
Non-trainable params: 3,904
```

EXAMPLE 1:
Approximating Least squares Reverse Time migration combining RTM and UNET in the image space

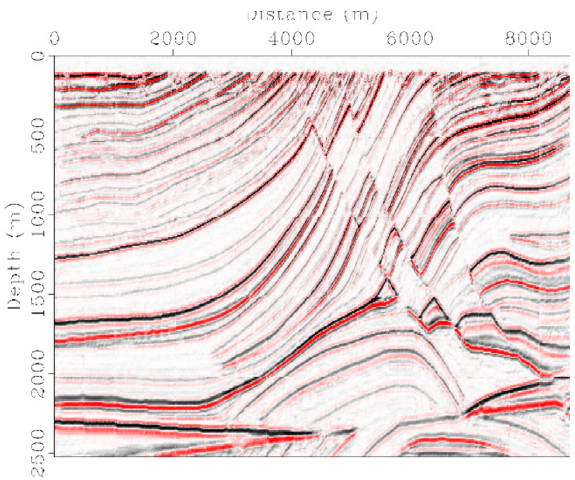
The problem: take regular migration and convert it to the ideal case



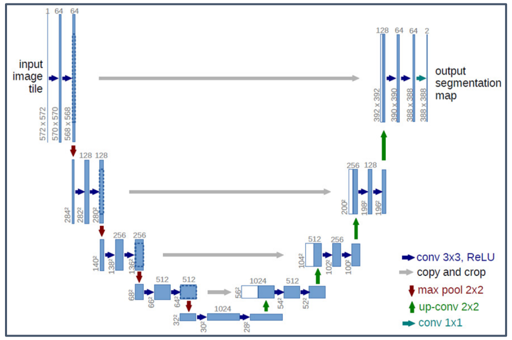
Regular RTM



Reflectivity from background velocity



Enhanced RTM

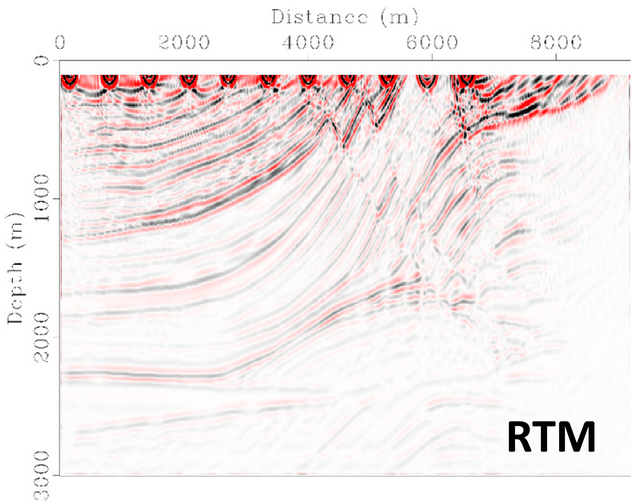


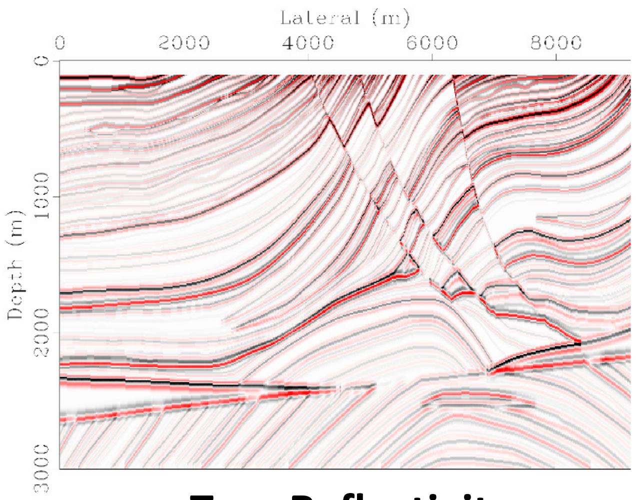
The problem: Regular RTM to ideal RTM

The questions: Labels? Data (channels)? Network?

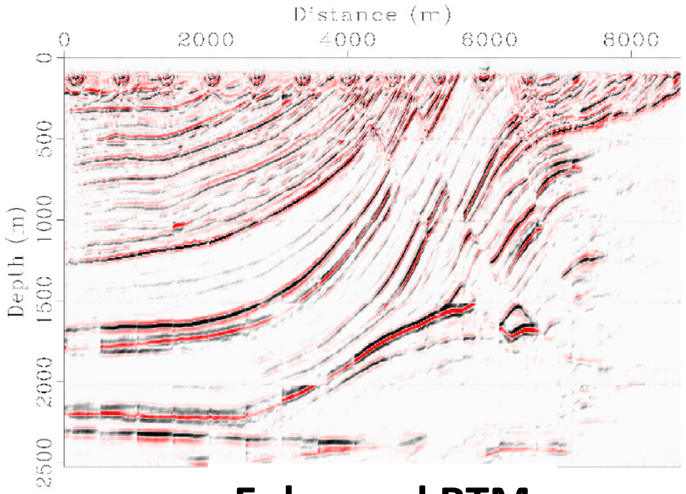
The challenge: Generalize?

RTM Option 1



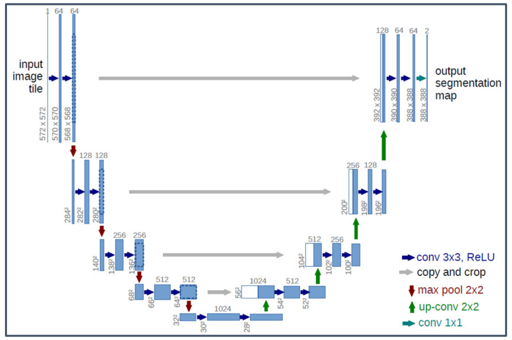


True Reflectivity

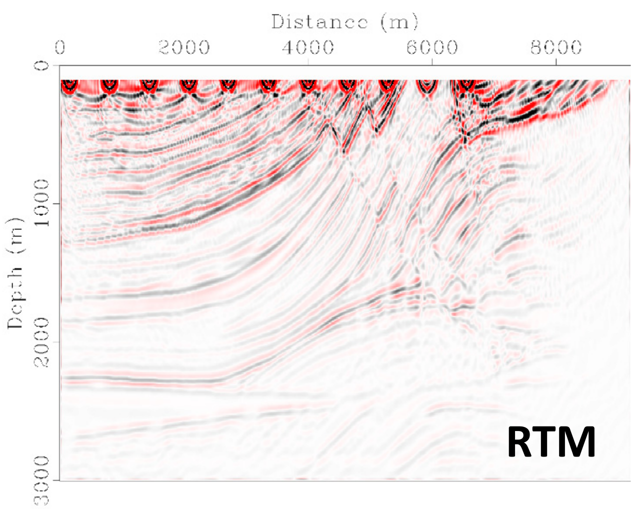


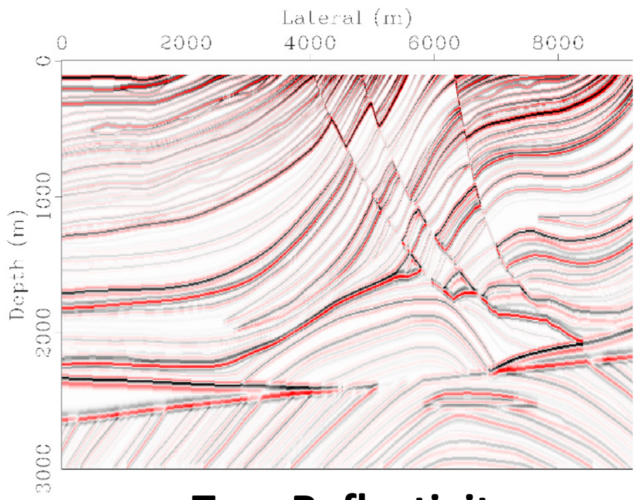
Enhanced RTM

(Training + prediction)

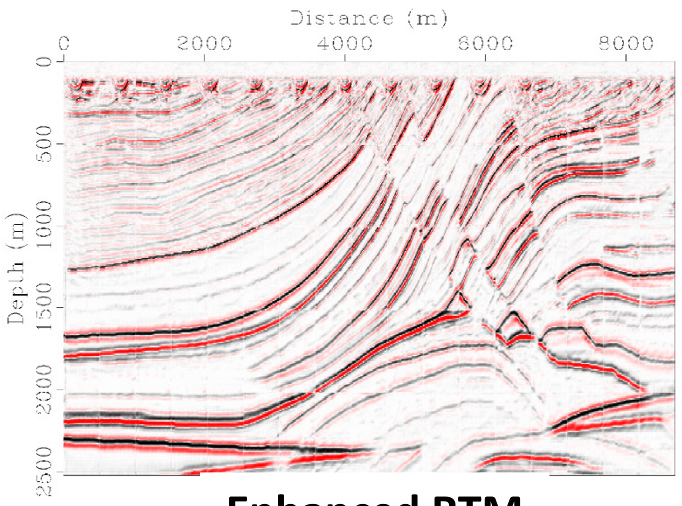


RTM Option 2

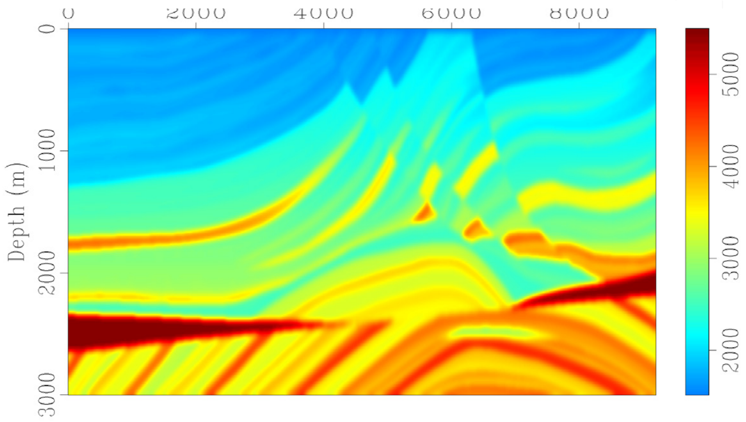




True Reflectivity

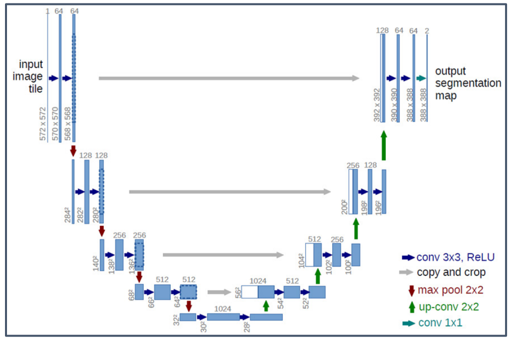


Enhanced RTM

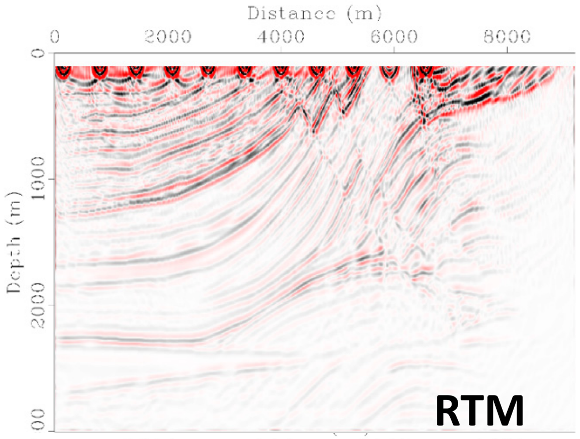


Background velocity

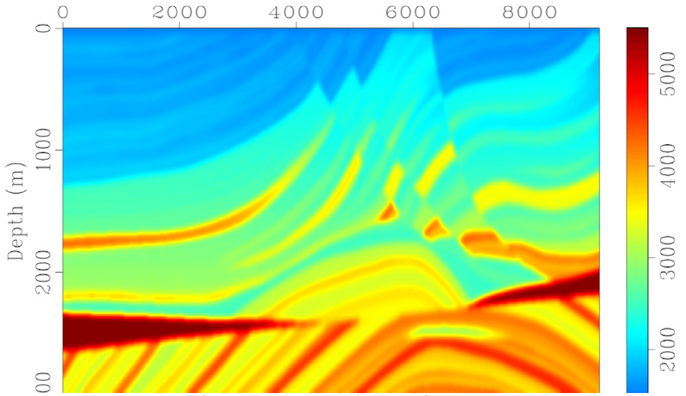
(Training + prediction)



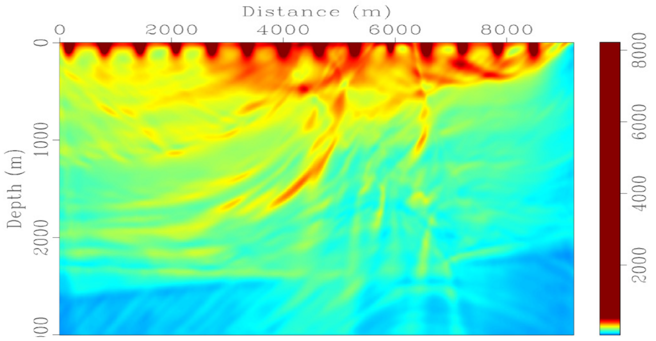
RTM Option 3



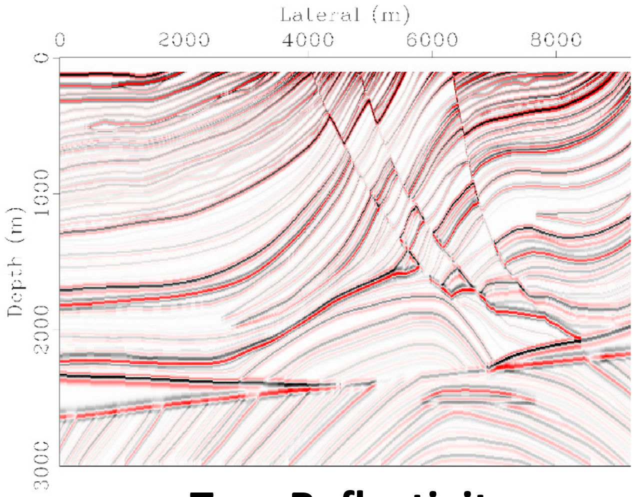
RTM



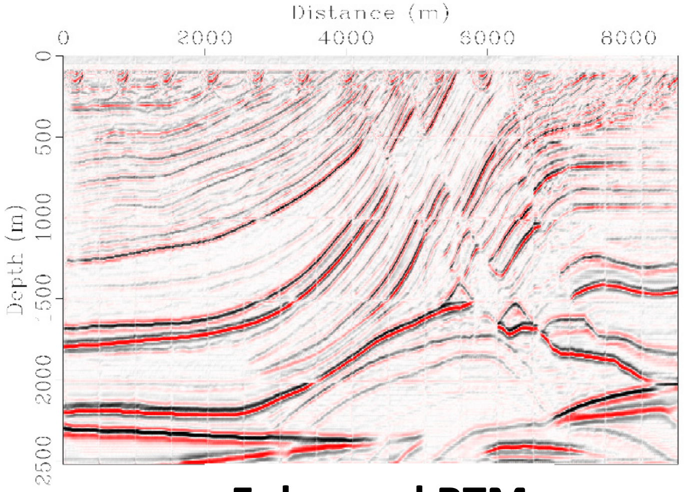
Background velocity



Illumination

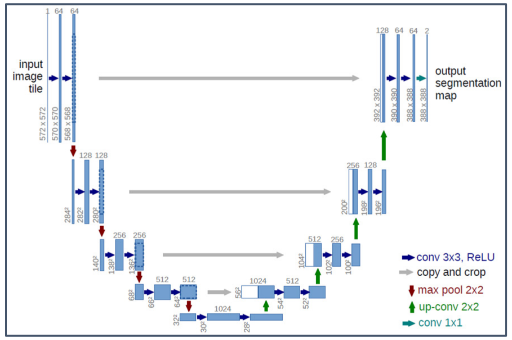


True Reflectivity

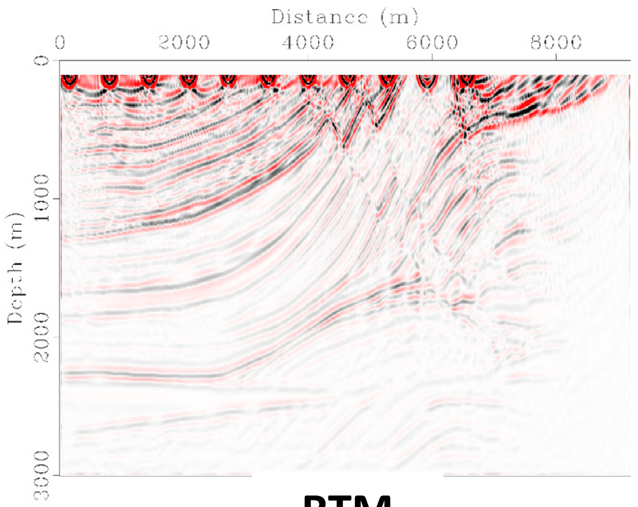


Enhanced RTM

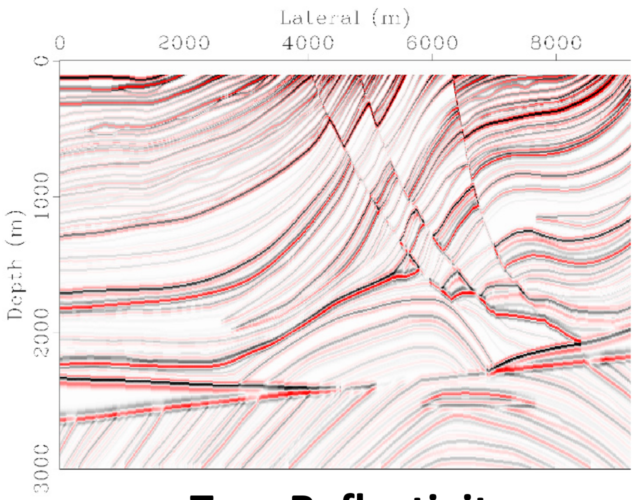
(Training + prediction)



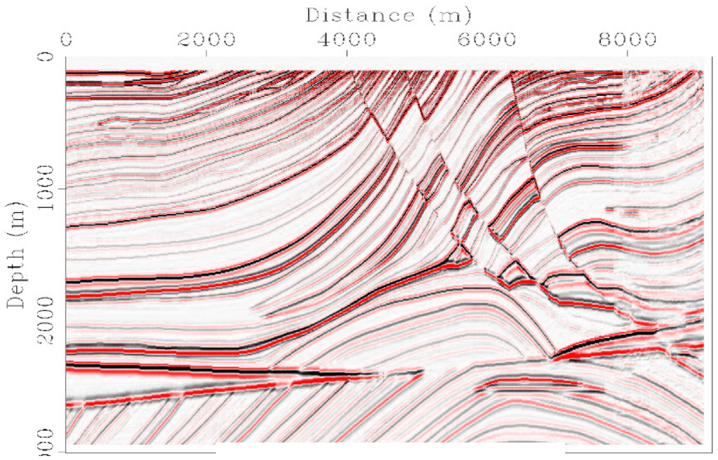
RTM option 4 (best)



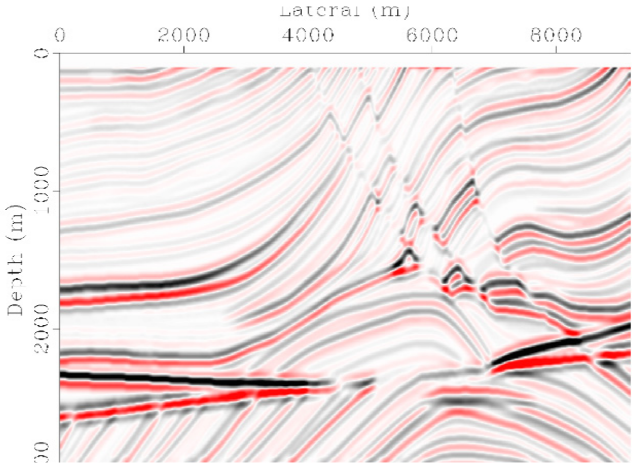
RTM



True Reflectivity

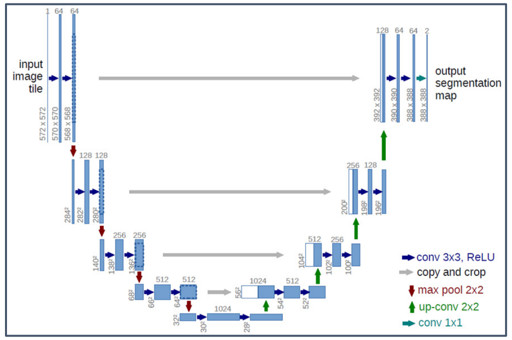


Enhanced RTM

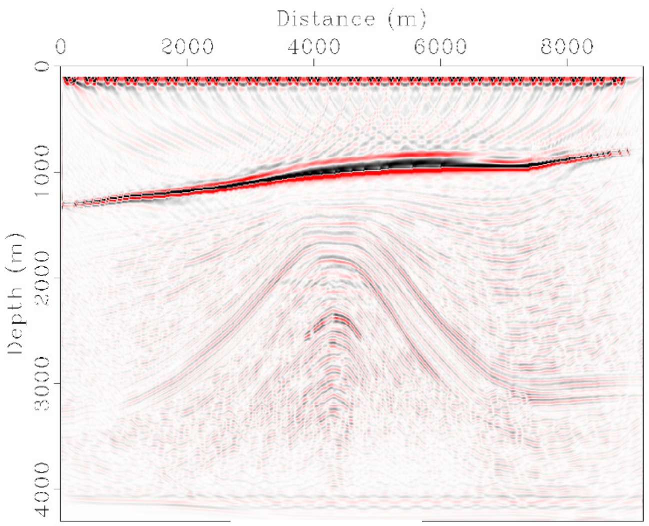


Reflectivity from background velocity

(Training + prediction)

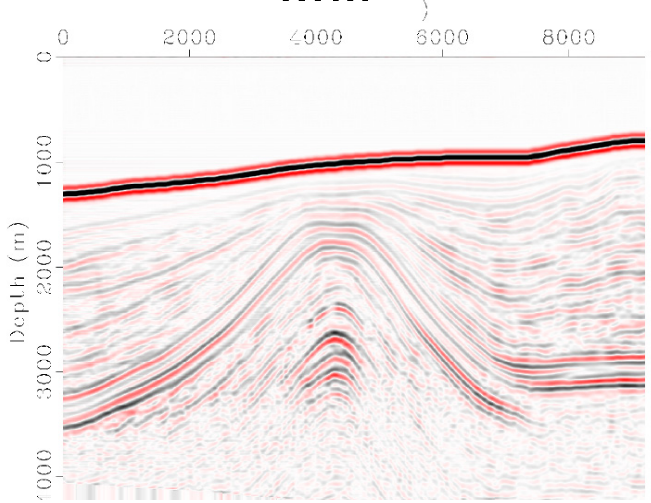


RTM for a model not used in training



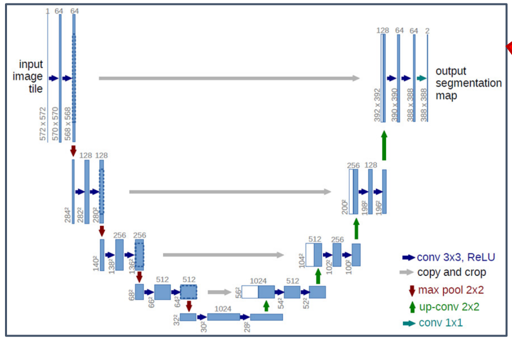
RTM

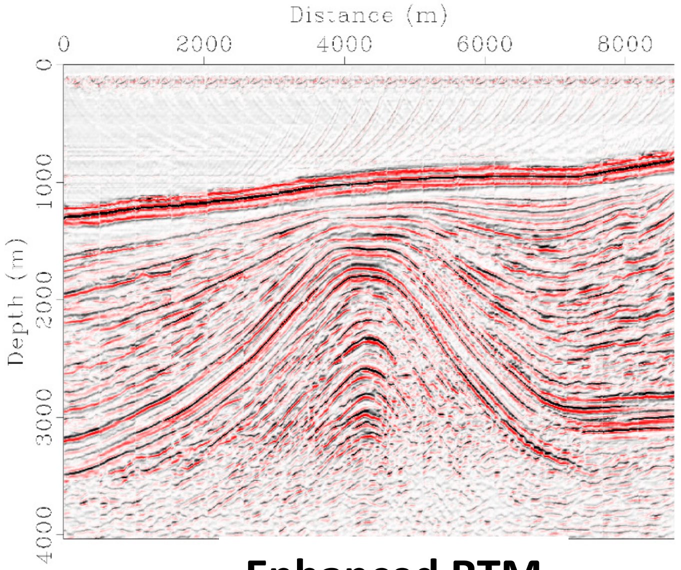
NO LABELS



Reflectivity from background velocity

(Prediction)





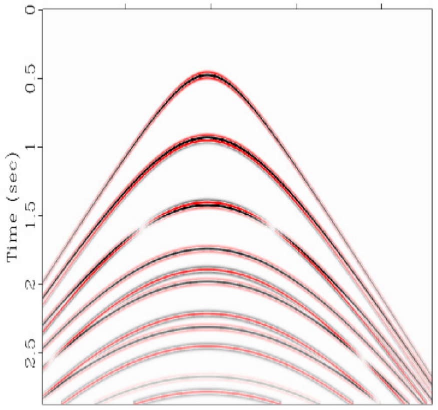
Enhanced RTM

EXAMPLE 2:
Attenuating multiples with combined Hyperbolic Radon and UNET mute

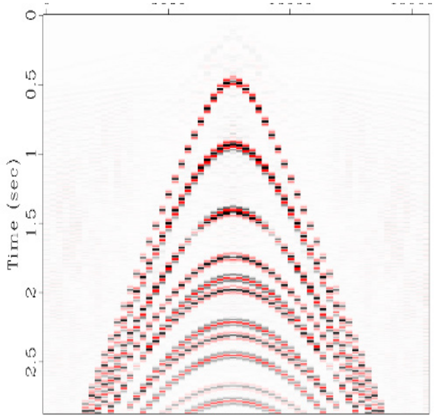
Example 2: Radon demultiple with ML-mute and Radon

Learn the mute filter in the Radon transform.
(675x200x1000 samples for data set split into 74000 windows 64x64x1)
Running 10 iterations with tensorflow-gpu (approximately 9 minutes).

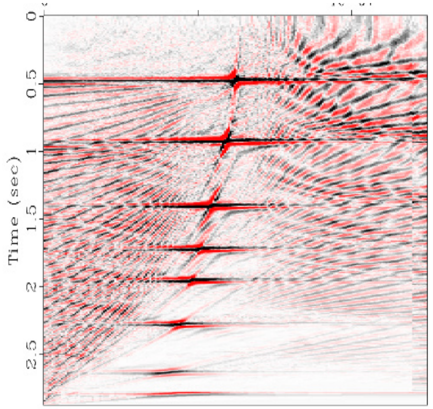
Shot



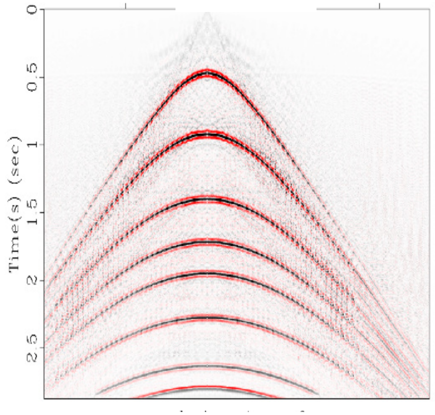
CMP



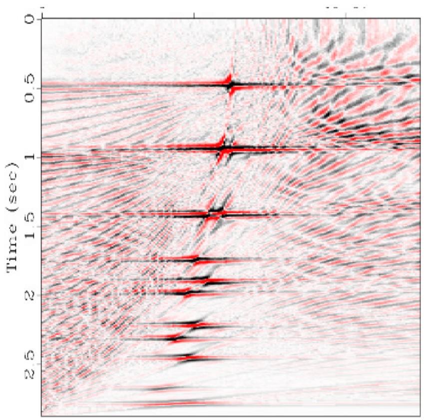
HRT

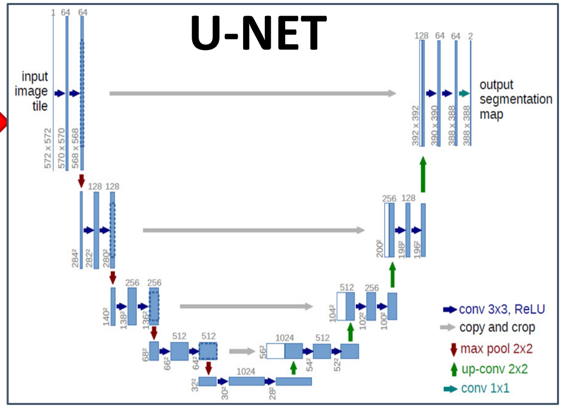


Shot



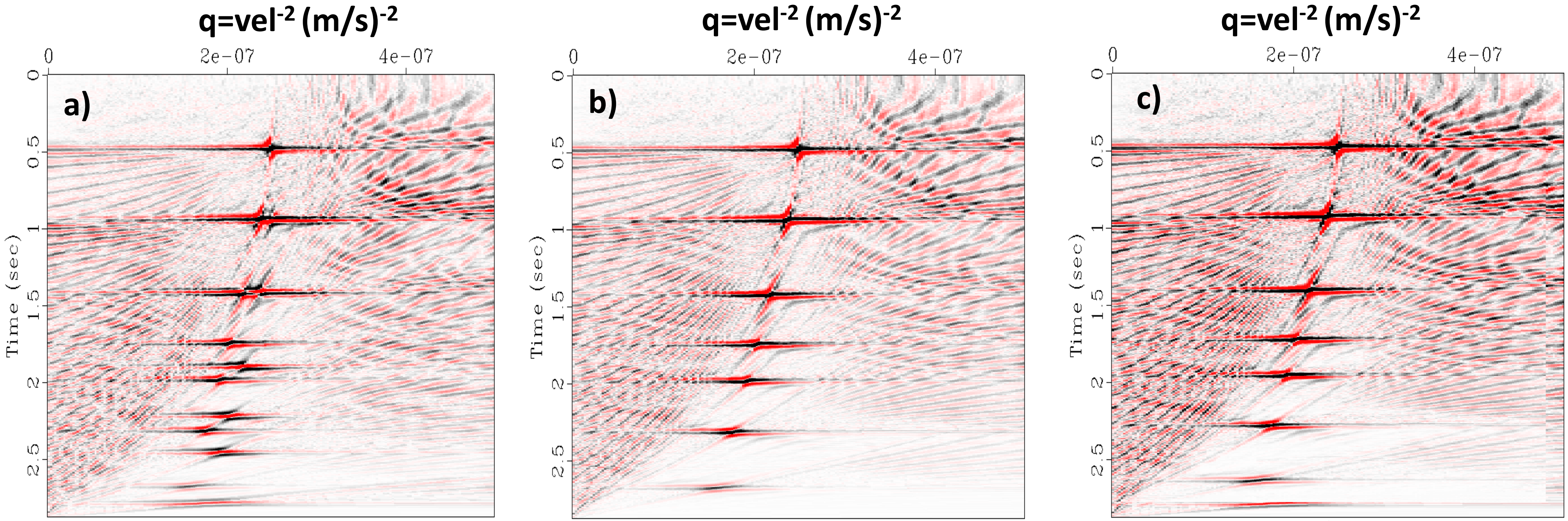
HRT





Example 2: Radon demultiple with Hyperbolic Radon

Calculate hyperbolic Radon transform in the CMP domain for both datasets.
(675 CMPs for each data set, 200 Radon parameters (1/Vel2) for each. (CUDA, 70 secs all cmps)

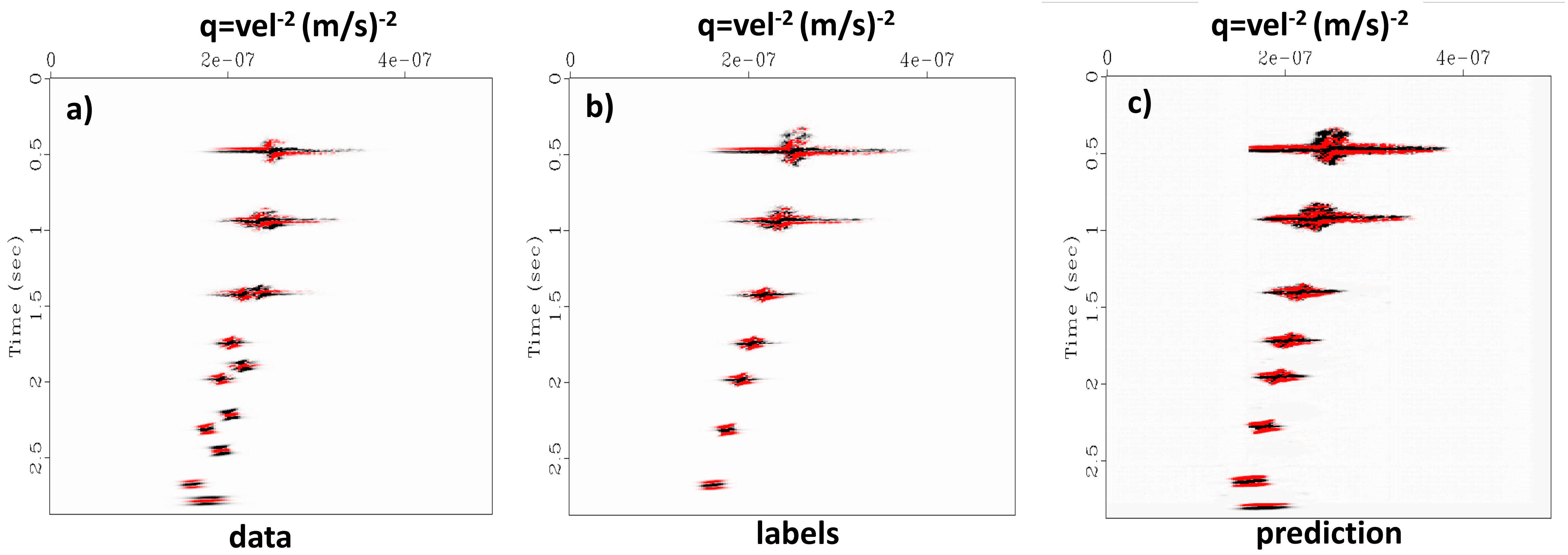


data labels prediction

Issue with resolution perhaps?

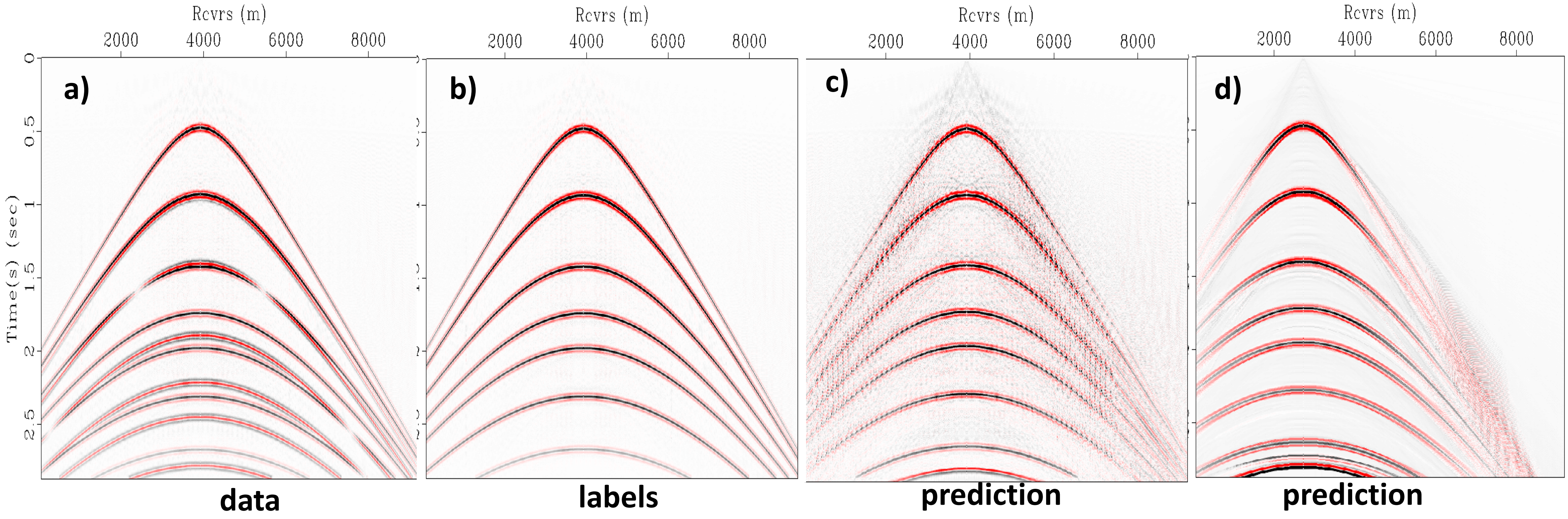
Example 2: Radon demultiple with Sparse Hyperbolic Radon

Calculate sparse hyperbolic Radon transform in the CMP domain for both datasets.
(675 CMPs for each data set, 200 Radon parameters (1/Vel2) for each.



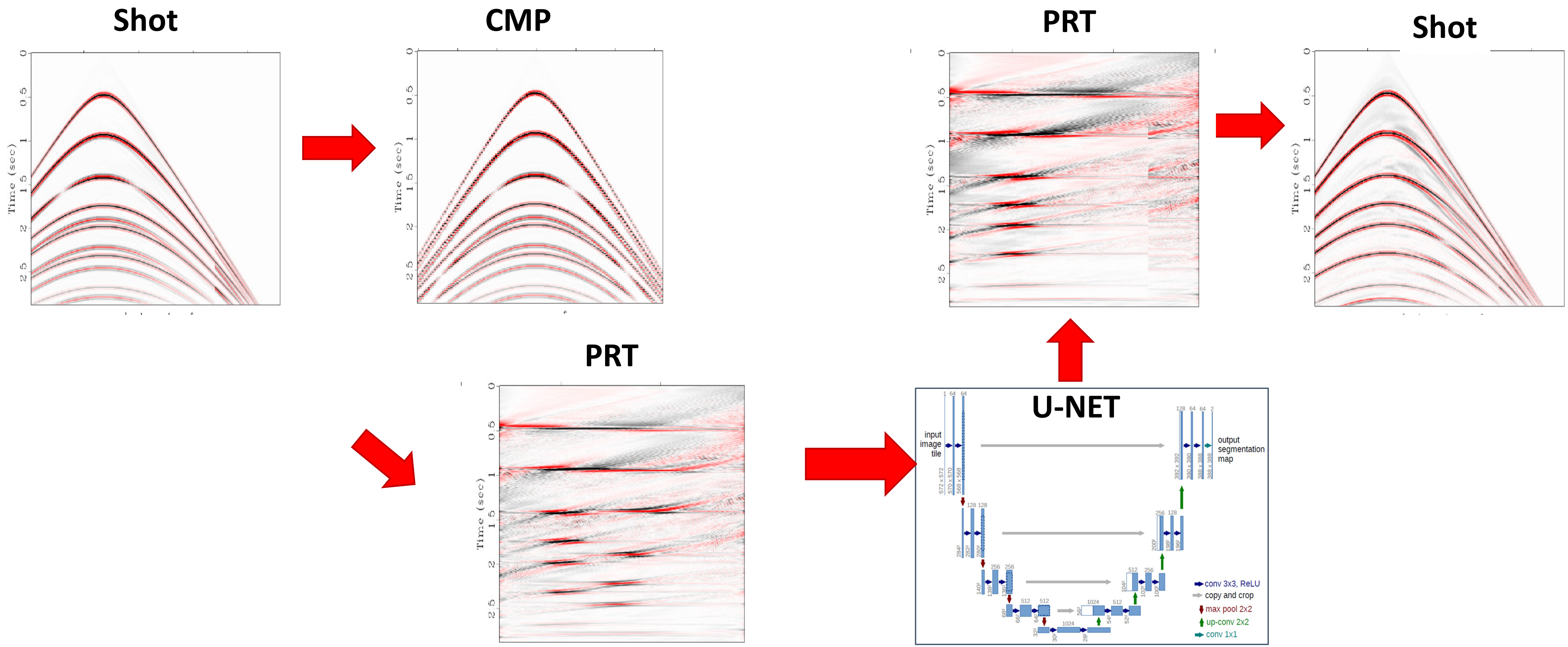
Would sparse Radon improve this?

Example 2: Increasing sampling

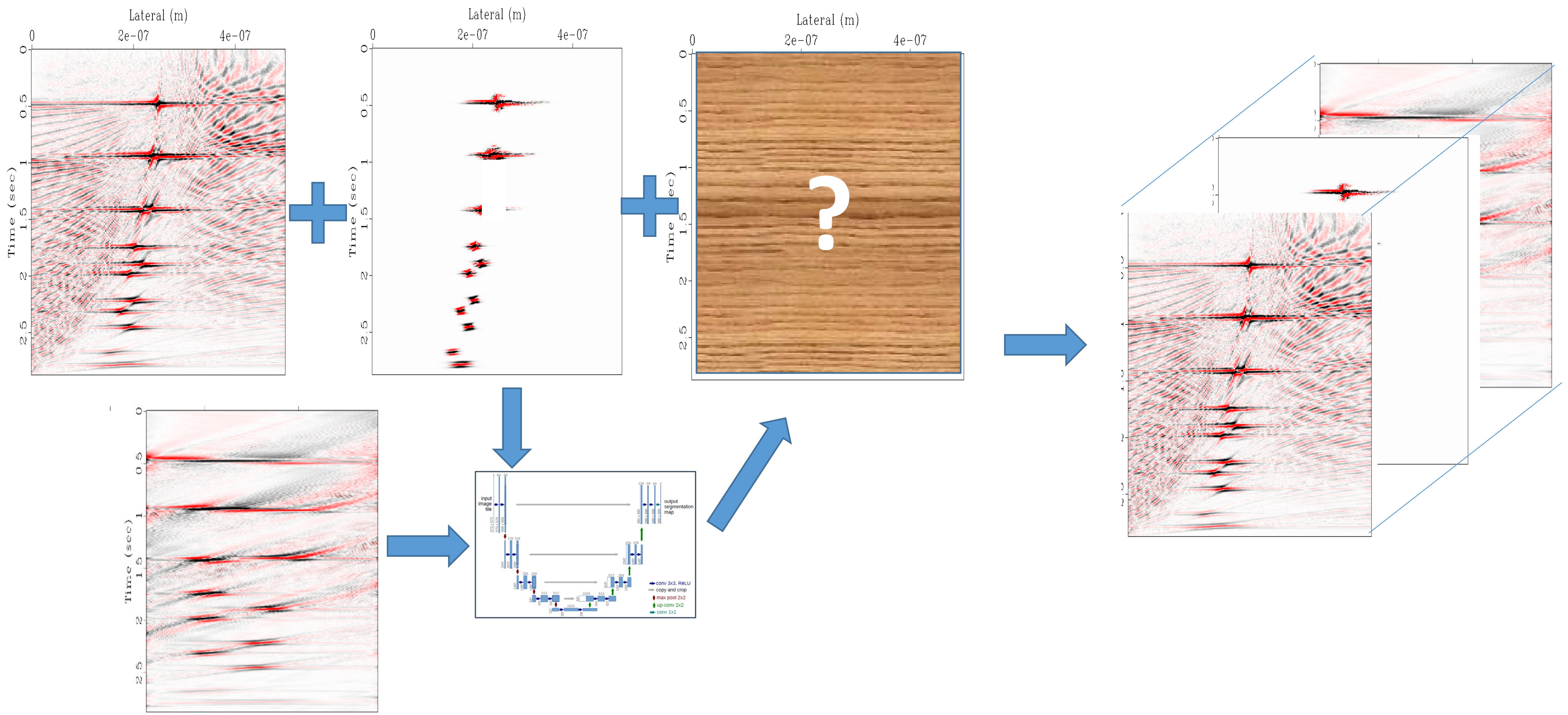


Example 2b: Radon demultiple with ML-mute and Parabolic Radon

Learn the mute filter in the Parabolic Radon transform.
(578x241x720 samples for data set split into 74000 windows 64x64x1)
Running 10 iterations with tensorflow-gpu (approximately 9 minutes).



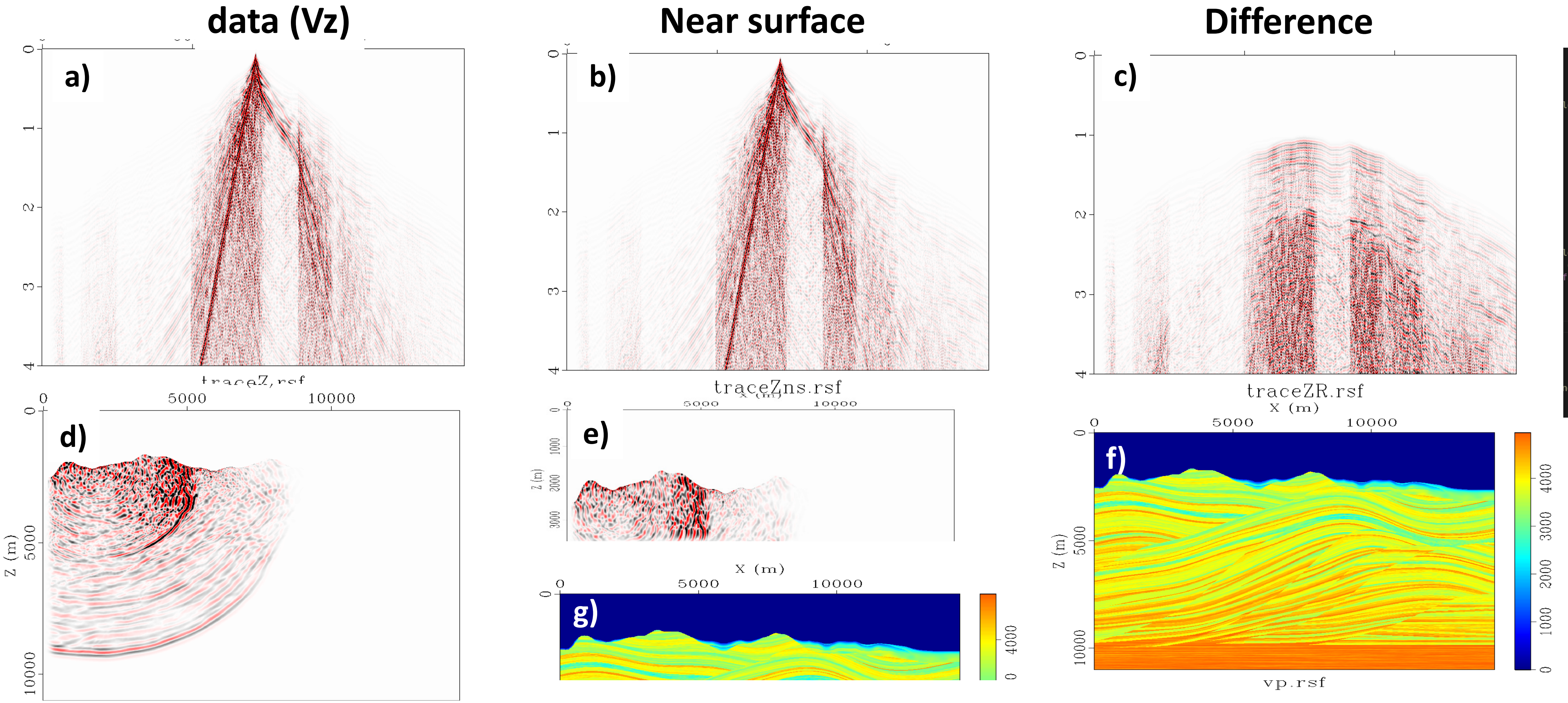
Combining information



EXAMPLE 3:
Attenuating ground roll with combined Finite difference modeling and UNET mute

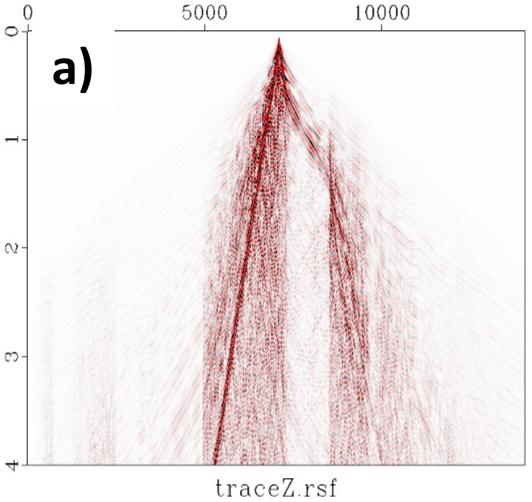
Example 3: Removing Ground Roll

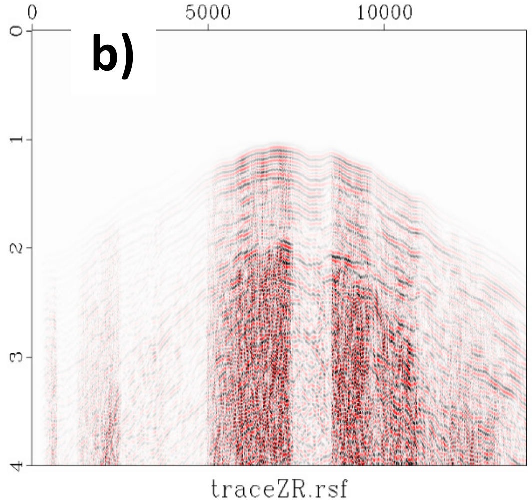
Finite difference elastic modeling from topography (program from Ivan Sanchez)
SEAM model (2D section). Predicting GR from near surface modeling

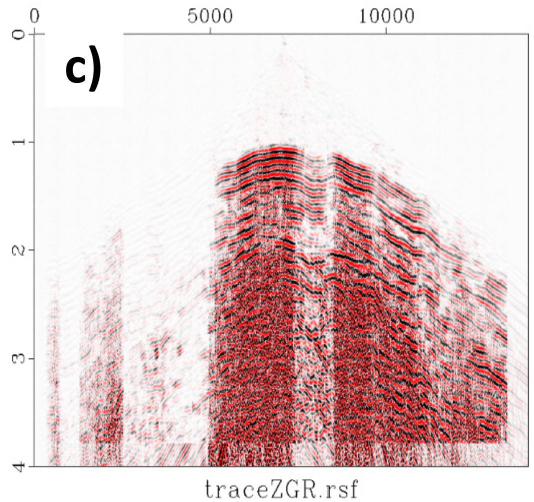


Example 3: Removing Ground Roll

Finite difference elastic modeling from topography (program from Ivan Sanchez)
Using Vp, Vs, Rho for SEAM model (2D section)





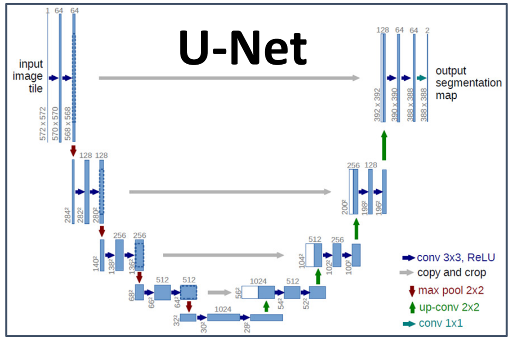


d)

?

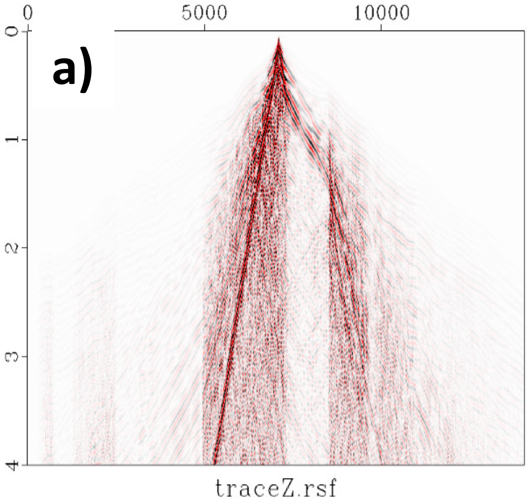
e)

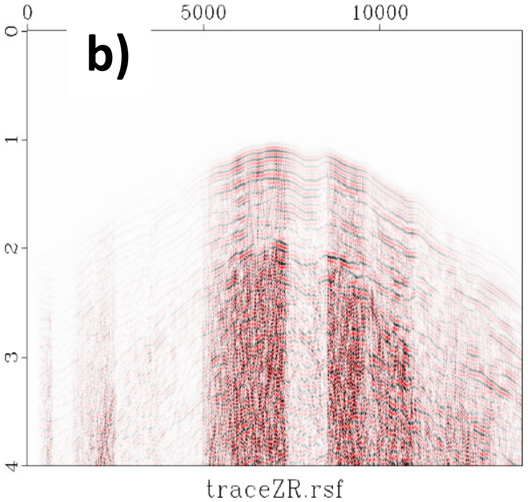
?

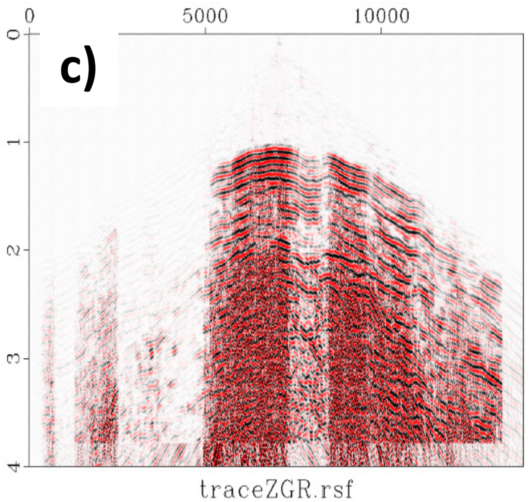


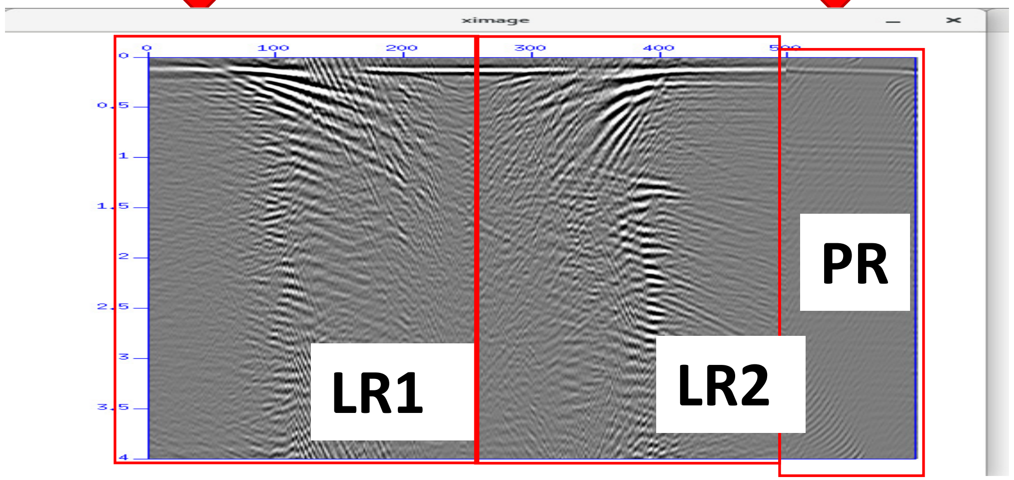
Example 3: Removing Ground Roll

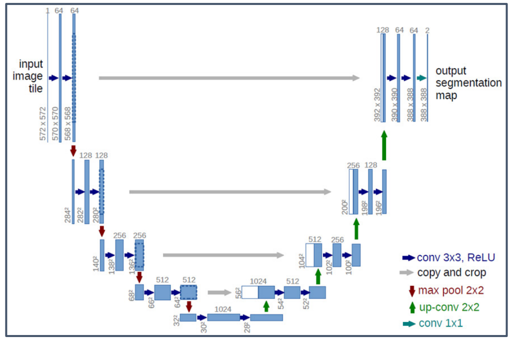
Finite difference elastic modeling from topography (program from Ivan Sanchez)
Using Vp, Vs, Rho for SEAM model (2D section)





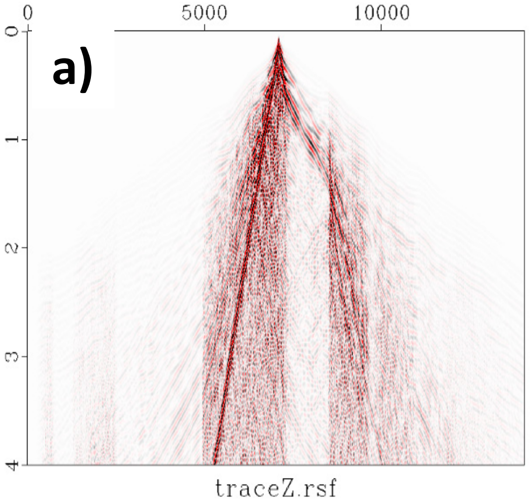


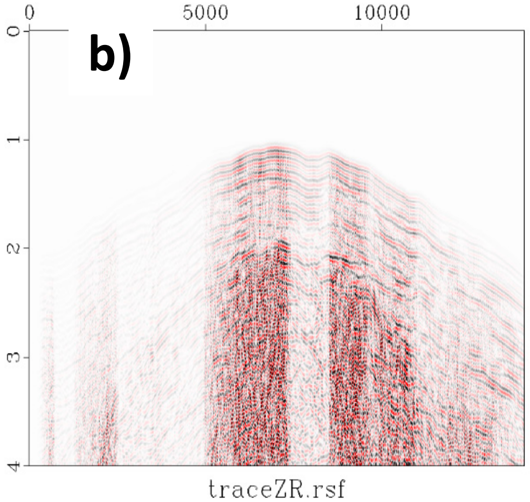


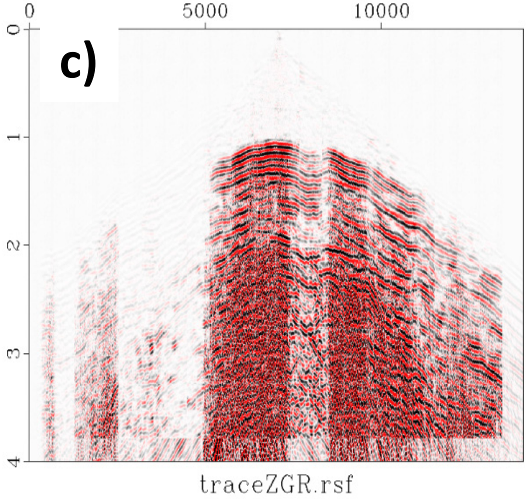


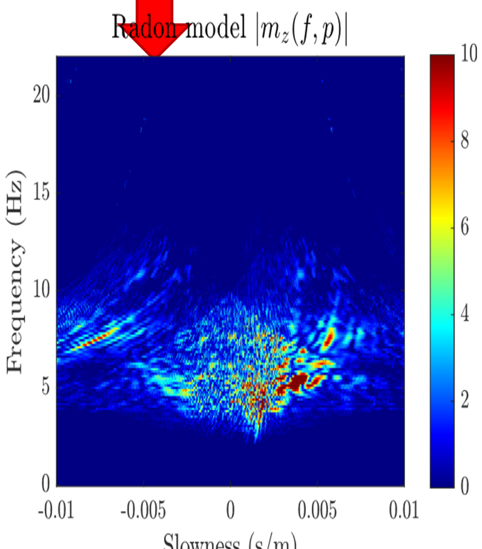
Example 3: Removing Ground Roll

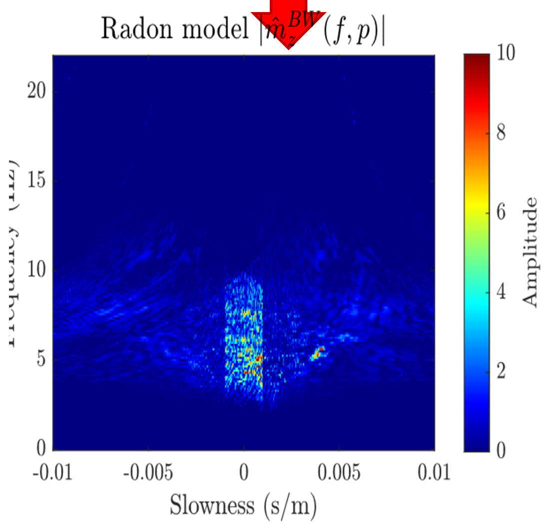
Finite difference elastic modeling from topography (program from Ivan Sanchez)
Using Vp, Vs, Rho for SEAM model (2D section)

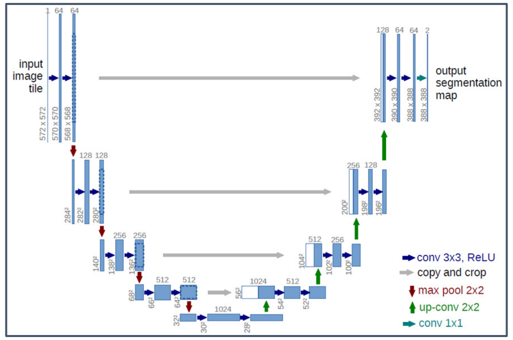






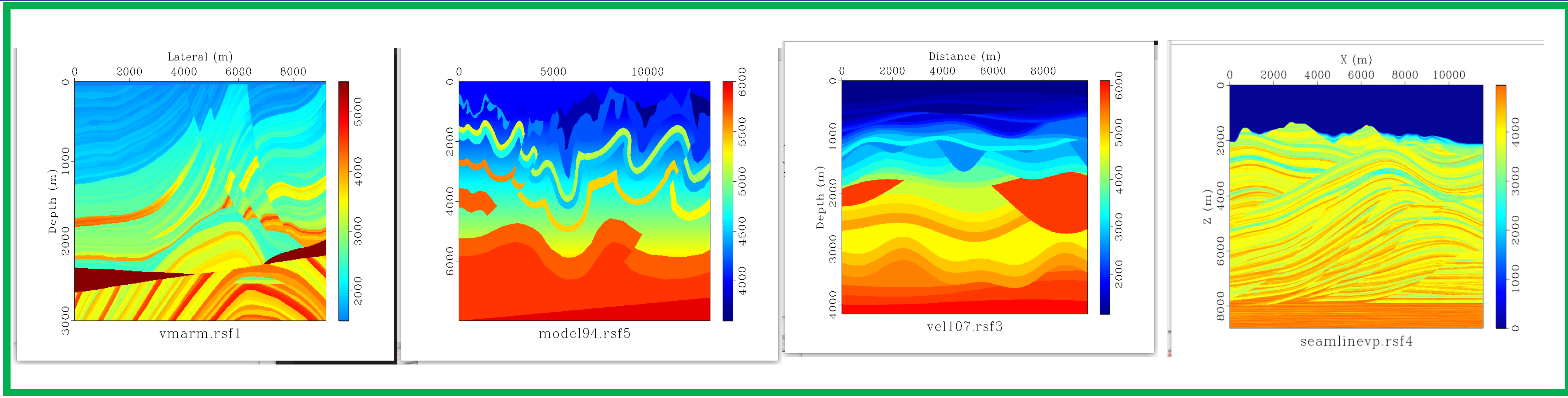




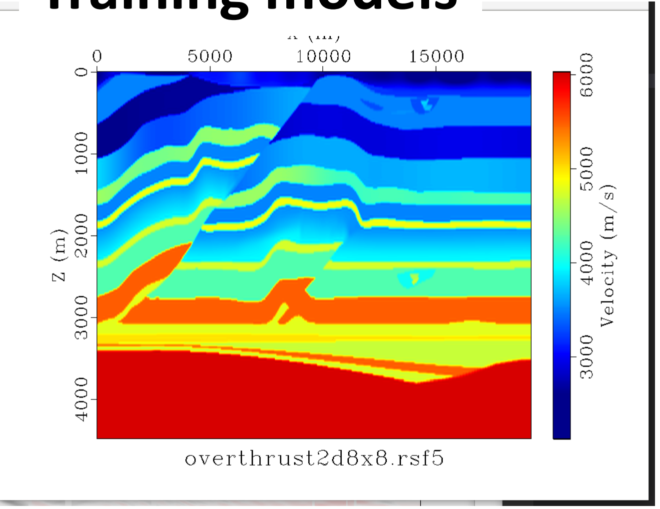


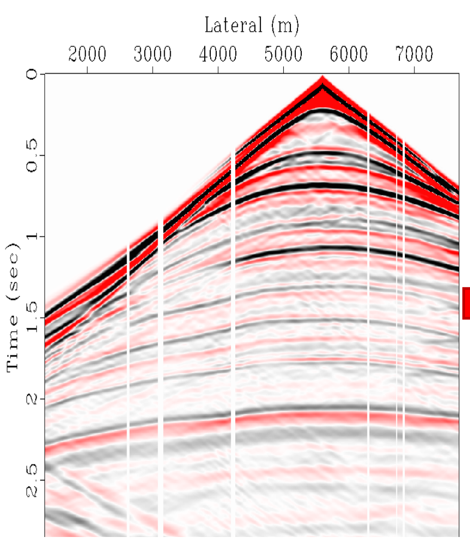
EXAMPLE 4:
Interpolation

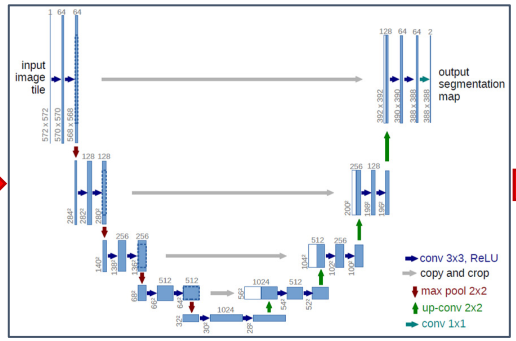
Interpolation

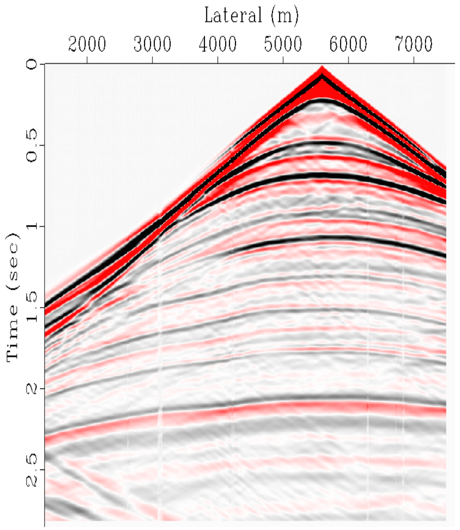


Training models



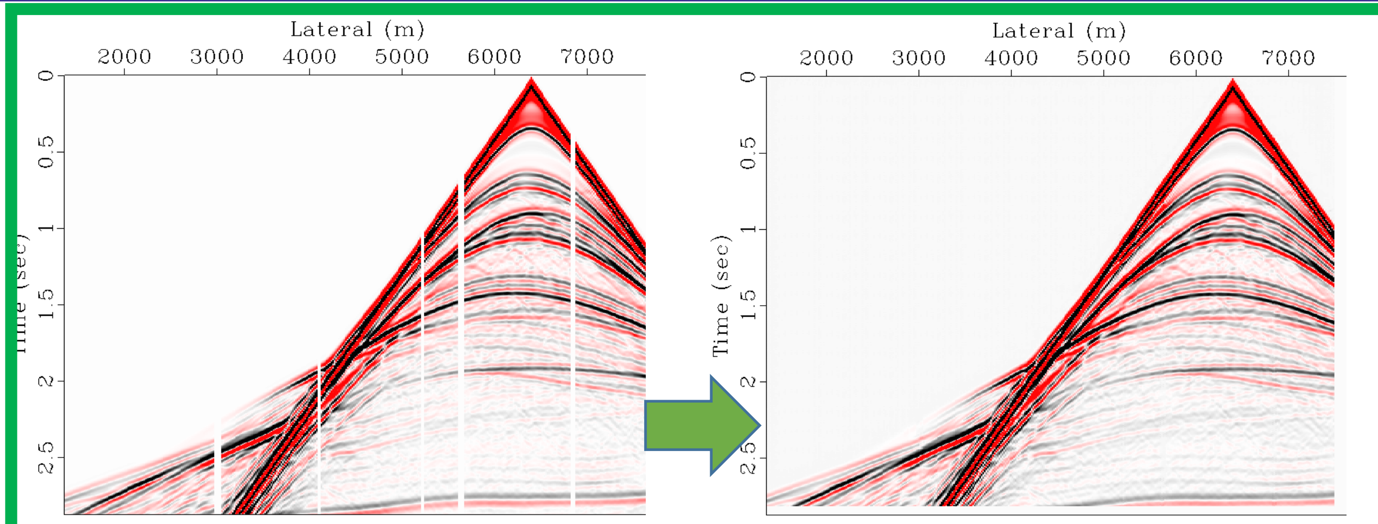




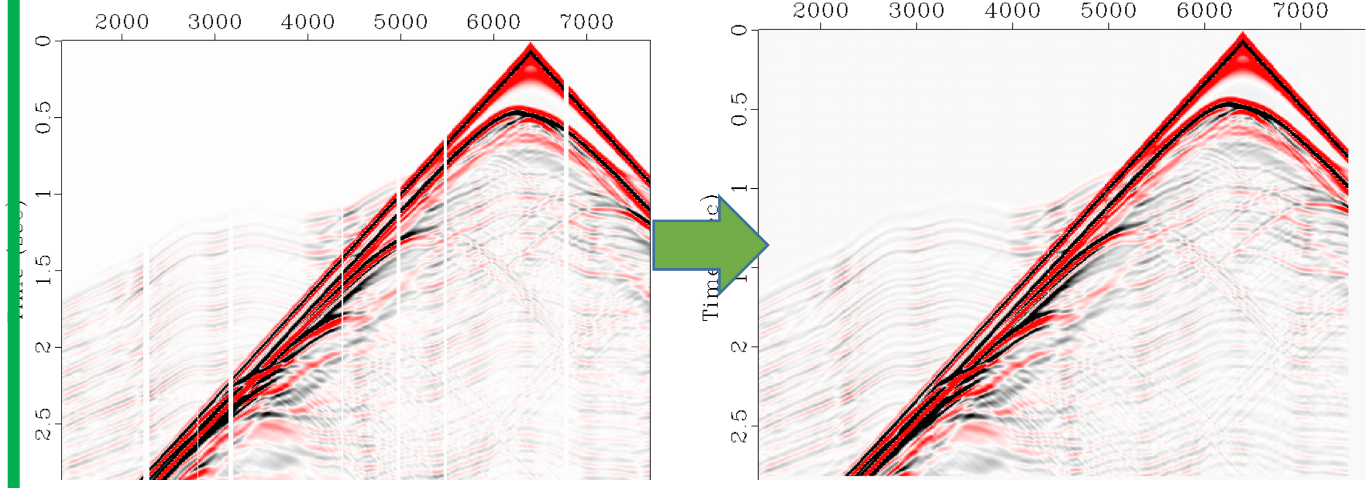


Testing model

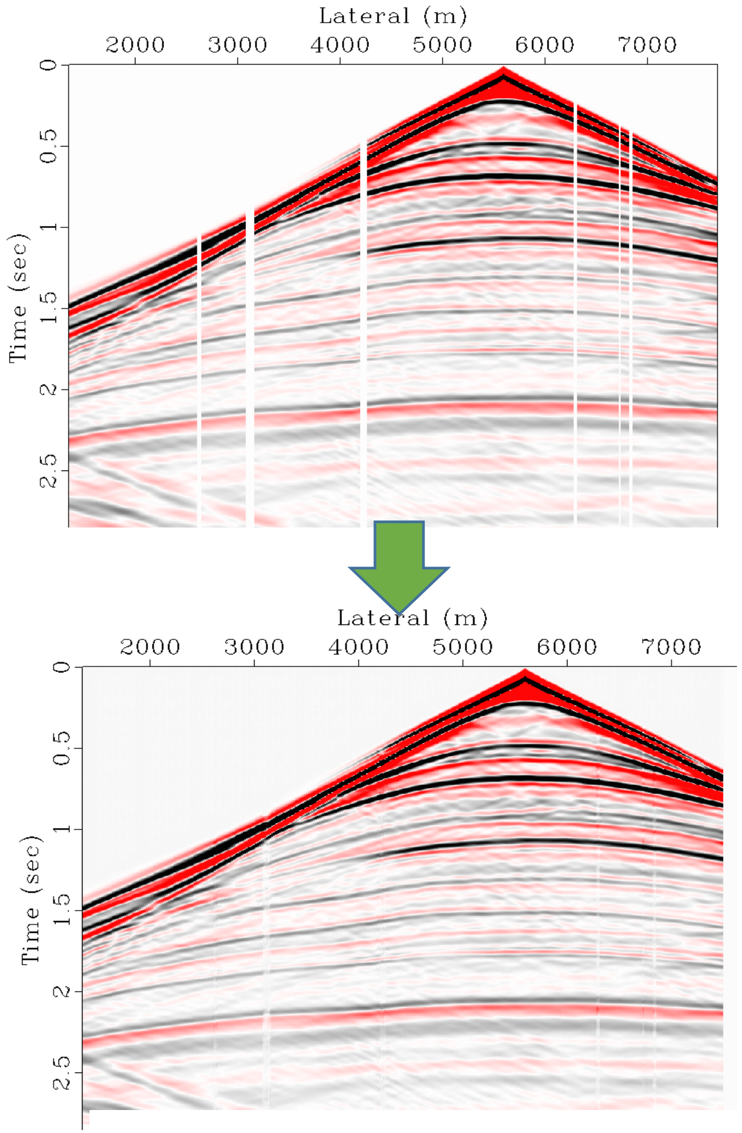
Interpolation



Random gaps



Prediction with training data



prediction with test data

Summary

1 – Information can be combined by adding different channels

2 – This information has to be in principle pixel to pixel equivalent but ..

3 – It maybe possible to combine different information through an additional network

4 – Different transformations can be used to make the DL job easier

5 – Different models can be combined to improve generalization.

Follow up

1 – Information can be combined: Shang Huang: LSRTM with multiples

2 – This information has to be in principle pixel to pixel equivalent but ..

3 – Use a network to combine information: Paloma Fontes: Demultiple with Radon

4 – Different transformations- Ivan Sanchez- GR attenuation with AE

5 – Different models can be combined to improve generalization: Daniel Trad

Ackowledgments

Support:

- CREWES sponsors
- Canadian SEG (Chair in Exploration Geophysics)
- Natural Sciences and Engineering Research Council of Canada

special thanks to:
Sam Gray (science and sanity check!)
Jorge Monsegny (figure out how to combine Madagascar with TF)
Ivan Sanchez (modeling ground roll)