Молодежная школа ННГУ

OpenVINO stories

Dmitry Kurtaev

Domains

- Retail
- Industrial
- Medical

Retail

Xiaogu Technology

- Facial recognition payment solution
- Android x86 OS
- Intel Apollo Lake N3350
  - CPU: Intel Atom® x7-E3950
  - GPU: Intel® HD Graphics 505

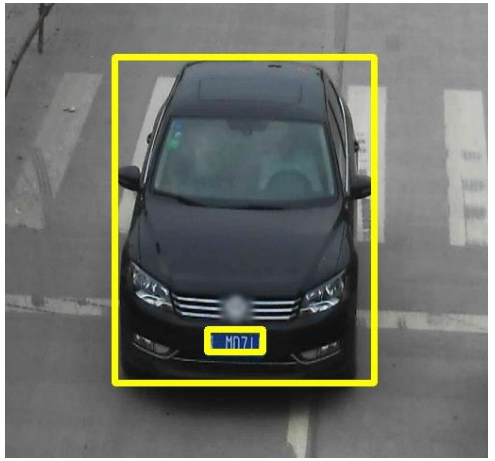
Algorithm	Execution time after acceleration with OpenVINO	
	CPU	GPU
Face detection	97 ms	48 ms
Face quality detection	40 ms	23 ms
Face liveness recognition	3 ms	2.5 ms
Face recognition	112 ms	57 ms

White paper: https://www.intel.cn/content/www/cn/zh/internet-of-things/xiaogu-3d-face-frictionless-payment-solution-en.html

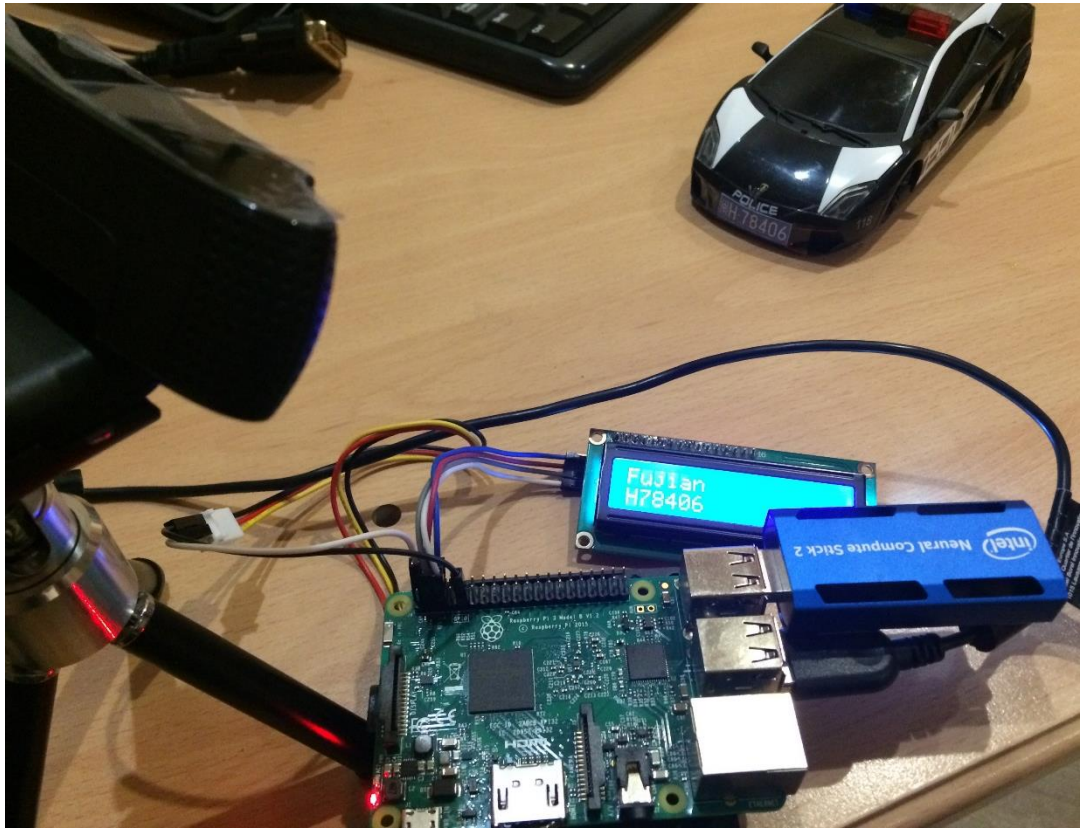
Xiaogu Technologies experience

- Do not implement classes in headers!
  - RTTI/Exceptions Not Working Across Library Boundaries
- cl_cache is powerful
- OpenCL for Android: Use Clang to build Clang to build Clang

Chinese License plate recognition

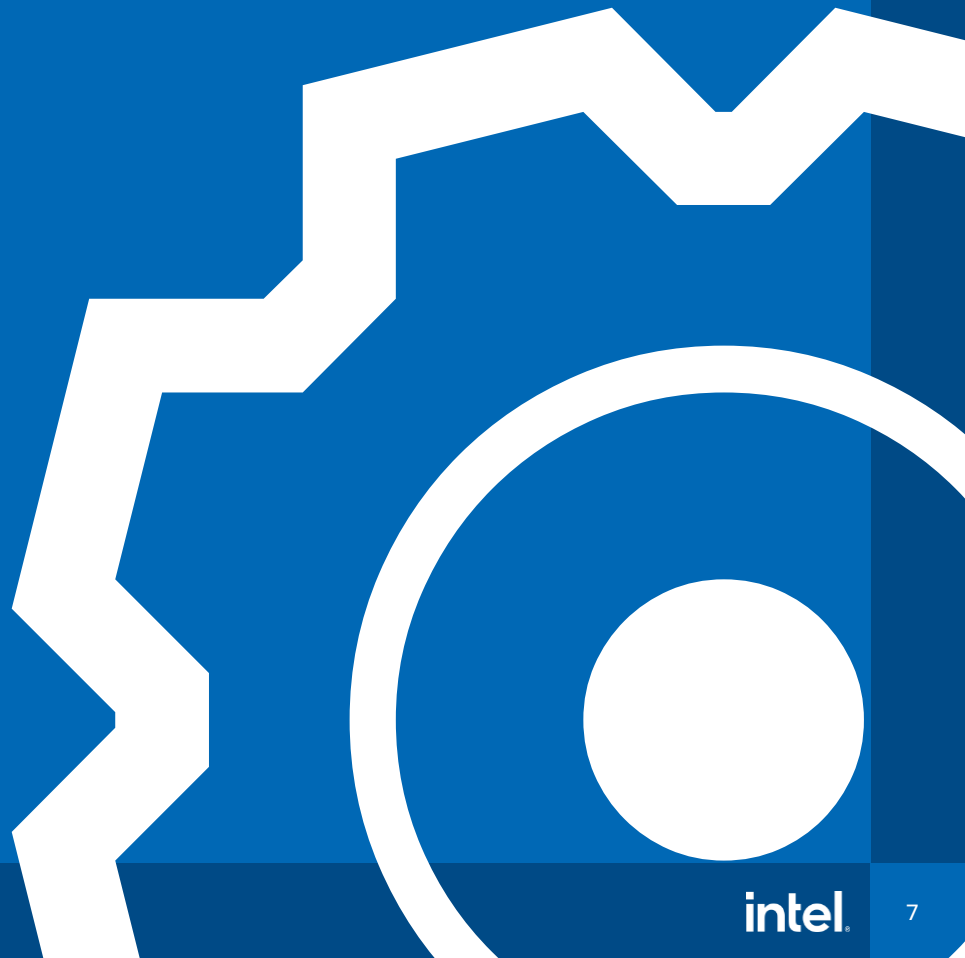


- 33 province letters
- regular, transit, public, military



[license-plate-recognition-barrier-0001]

Industrial

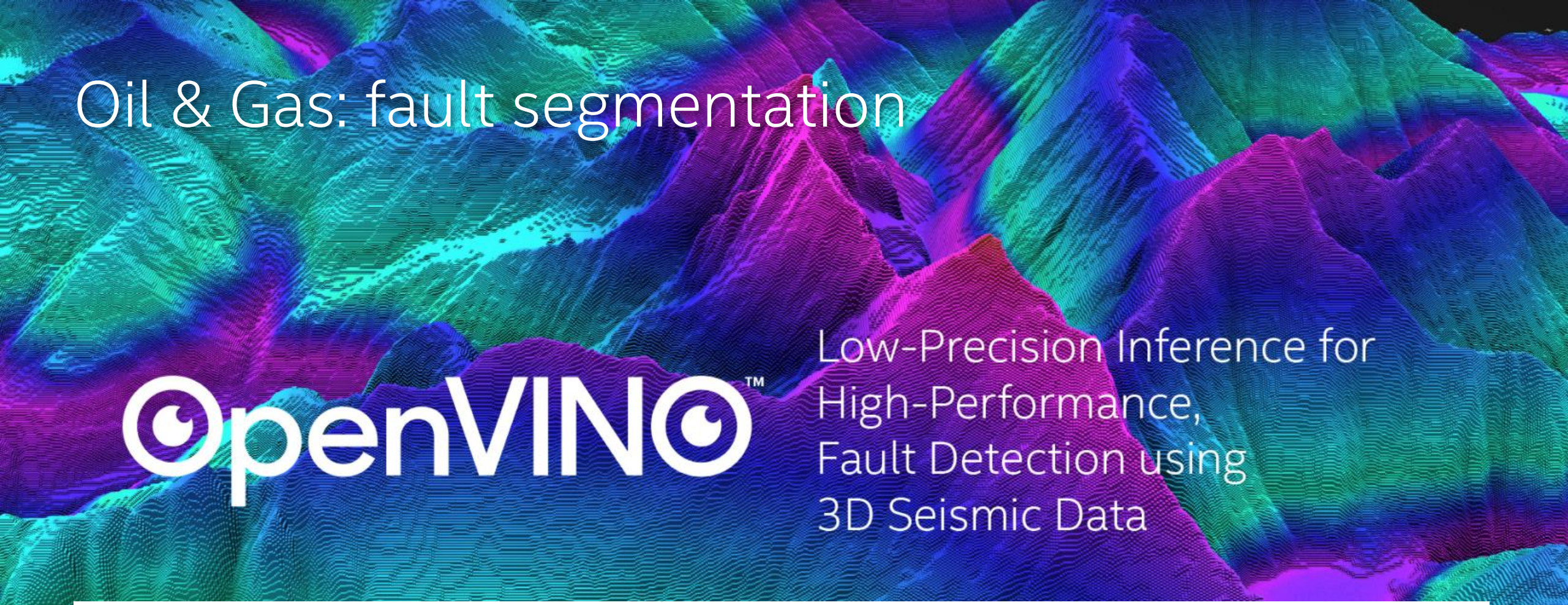


Weld porosity detection

- Robotic arc weld
- Human vision inspection leads to many false positives

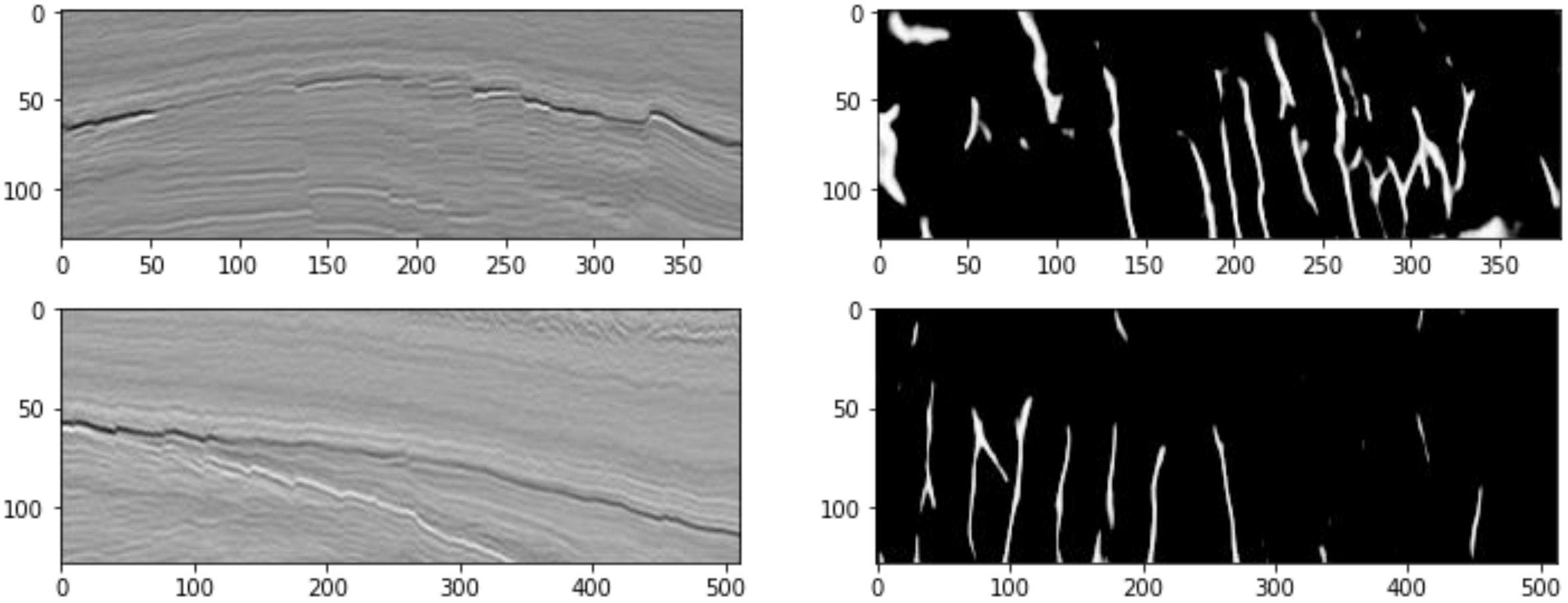
/openvinotoolkit/open_model_zoo/blob/master/models/intel/weld-porosity-detection-0001

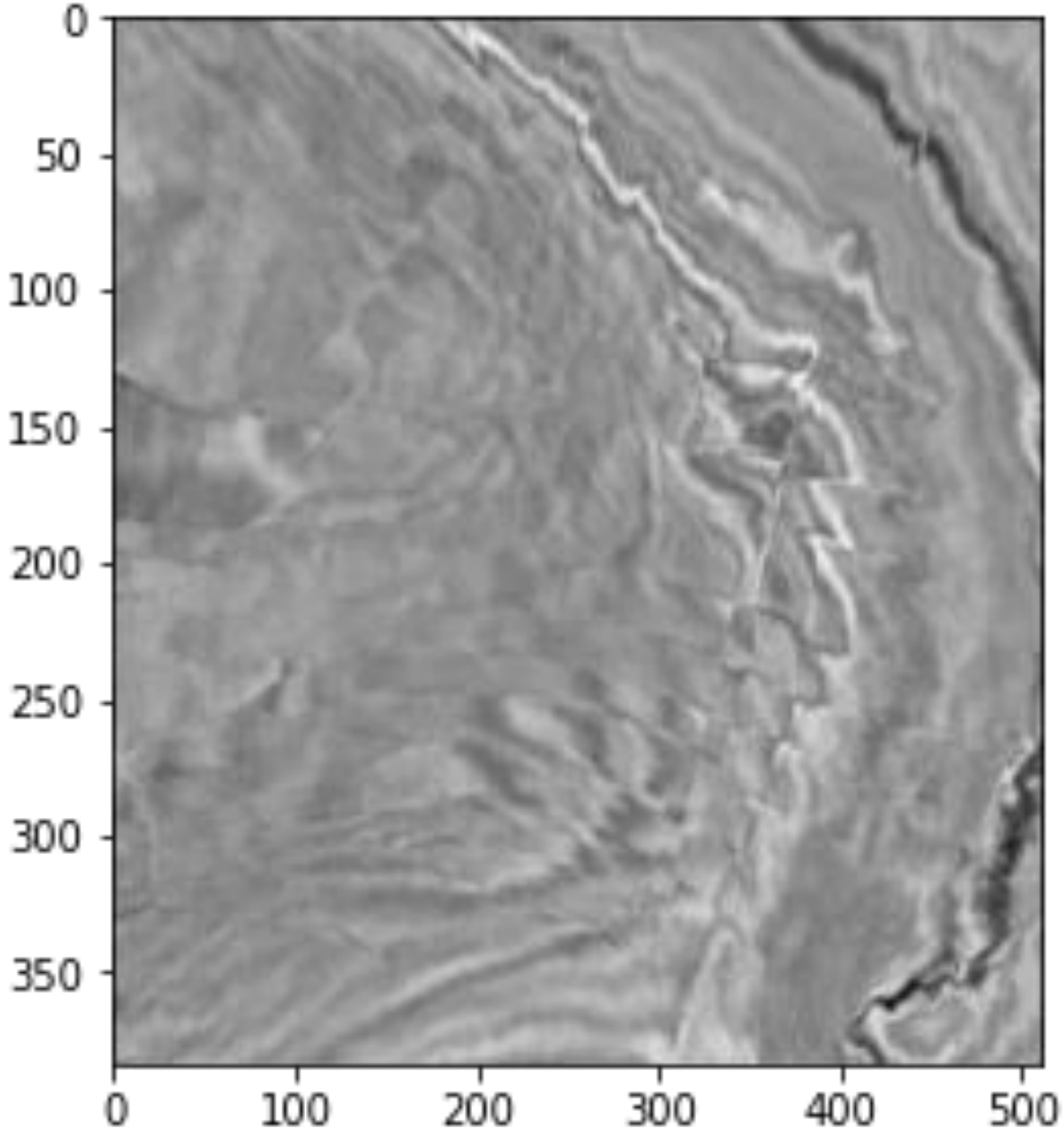
Oil & Gas: fault segmentation

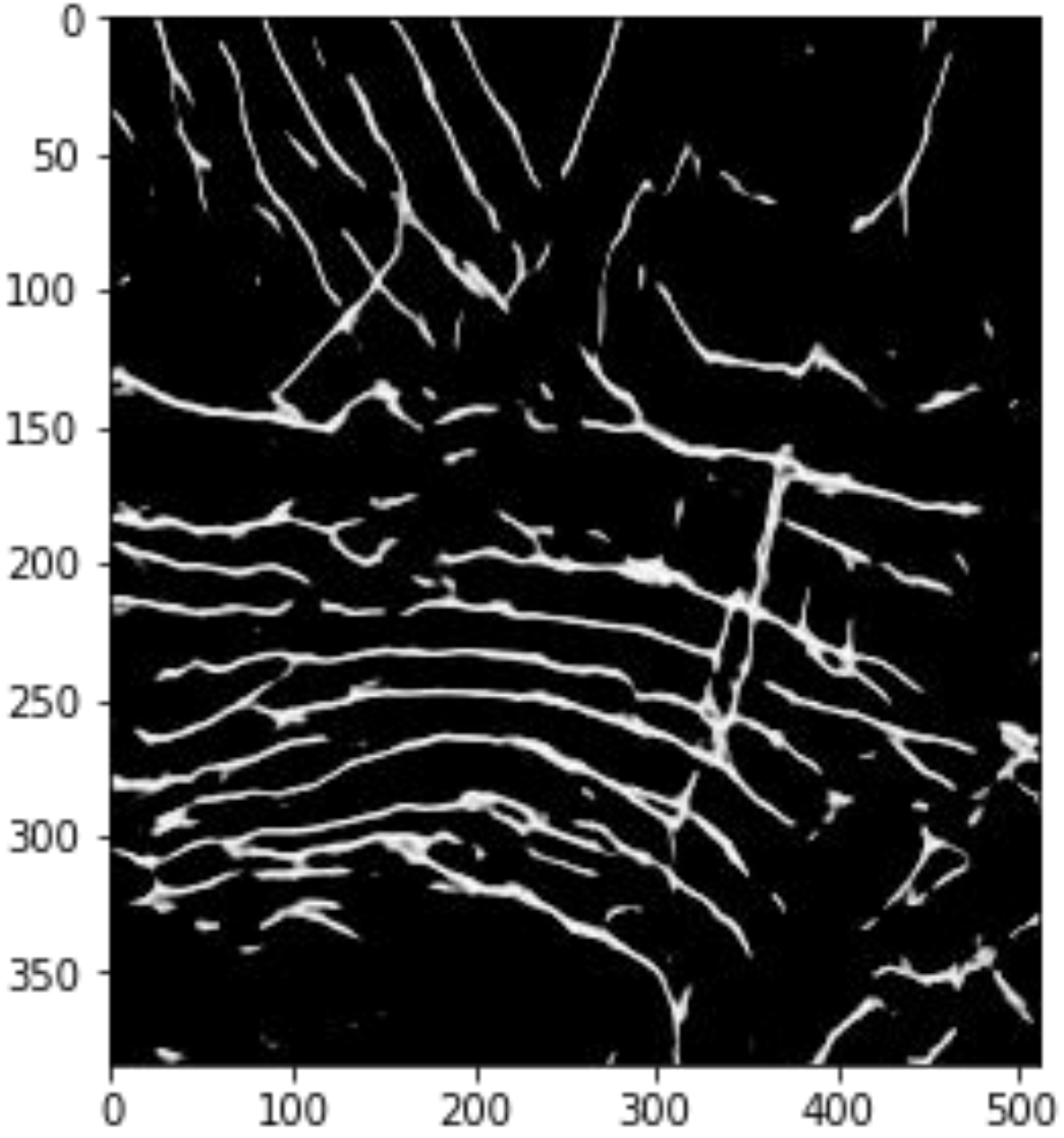


- Accelerating fault detection in 3D Seismic data using OpenVINO - Reducing time to the first Oil
- Analyzing 3D Seismic Data using Intel Distribution of OpenVINO Toolkit

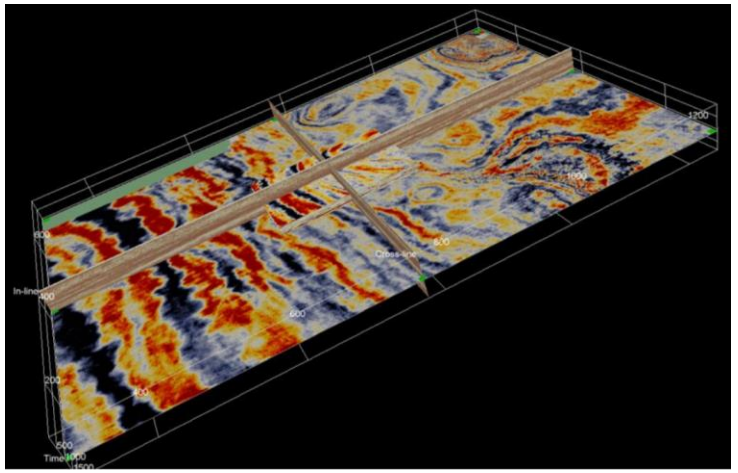
Oil & Gas: fault segmentation



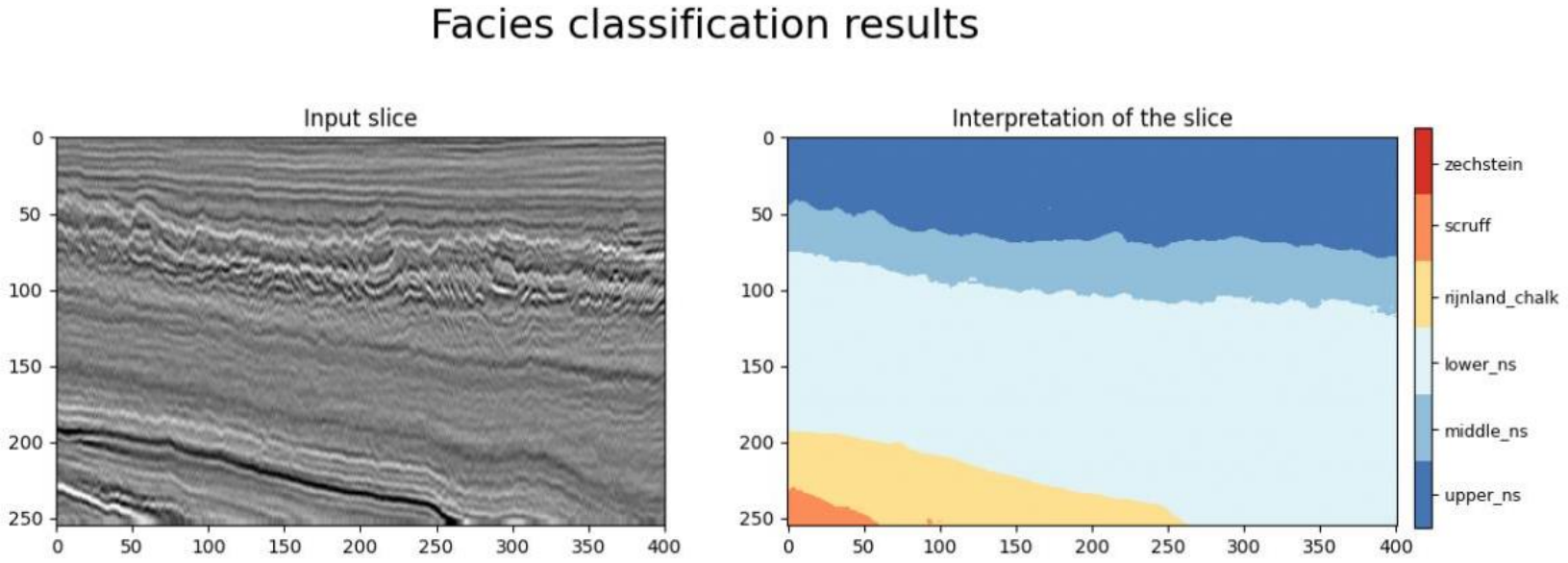

0
50
100
150
200
250
300
350
0
100
200
300
400
500

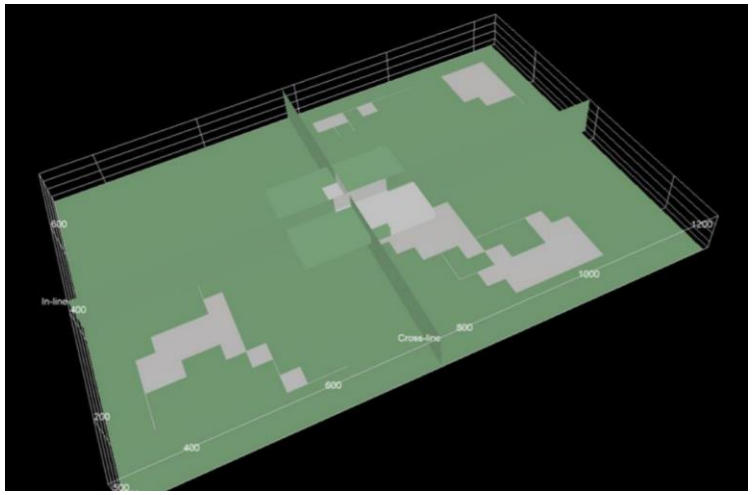

0
50
100
150
200
250
300
350
0
100
200
300
400
500

Oil & Gas: salt detection



Original 3D seismic section





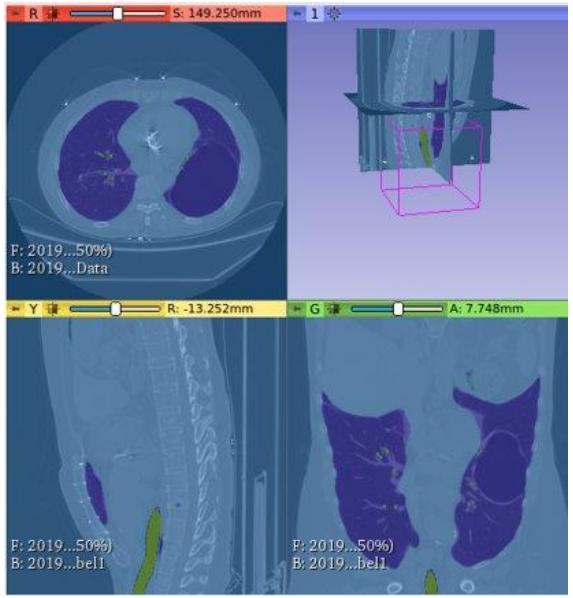
Salt prediction

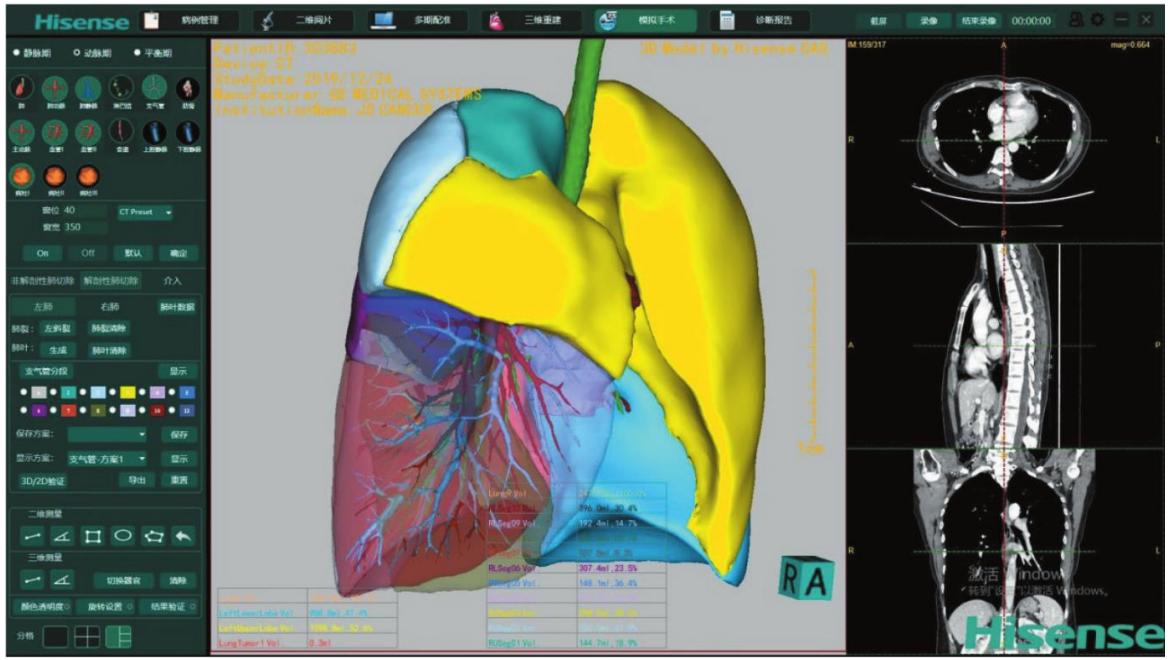
Medical

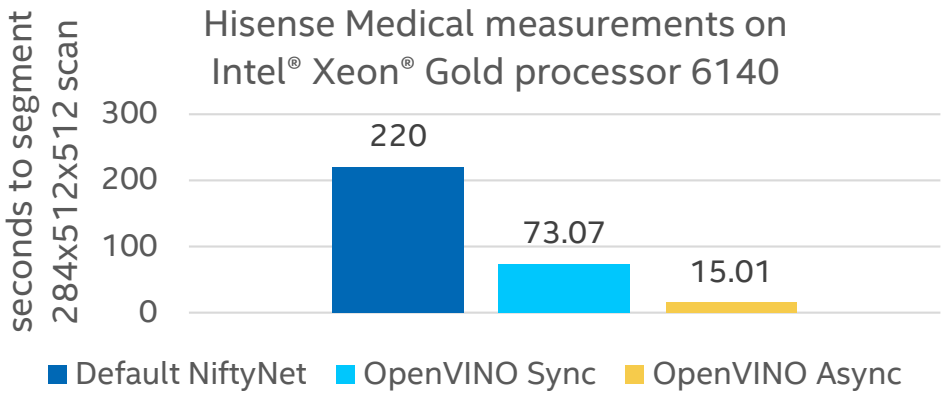
Surgery assistants

Hisense Medical

- Pulmonary surgery assistant
- NiftyNet framework based
- Dense VNet (3D UNet)





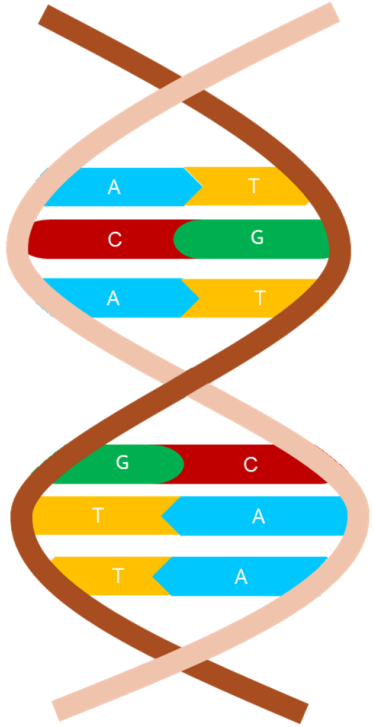


White paper: Hisense Medical Computer-assisted Surgery (CAS) System for Pulmonary Surgery based on Intel® Architecture

Genomics

Nucleotides

* Adenine
* Cytosine
* Guanine
* Thymine



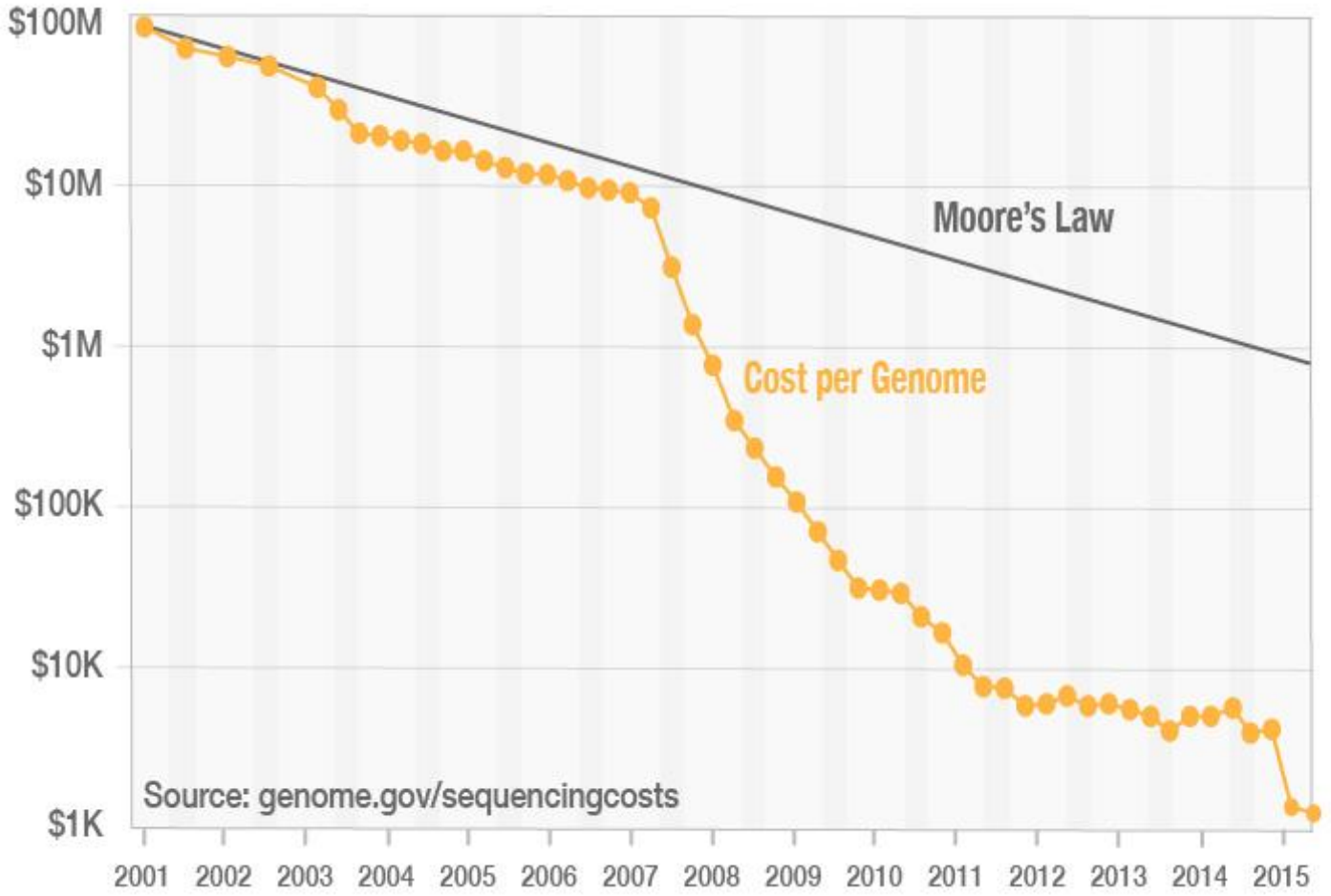
Base pairs

A - T

C - G

Sequencing

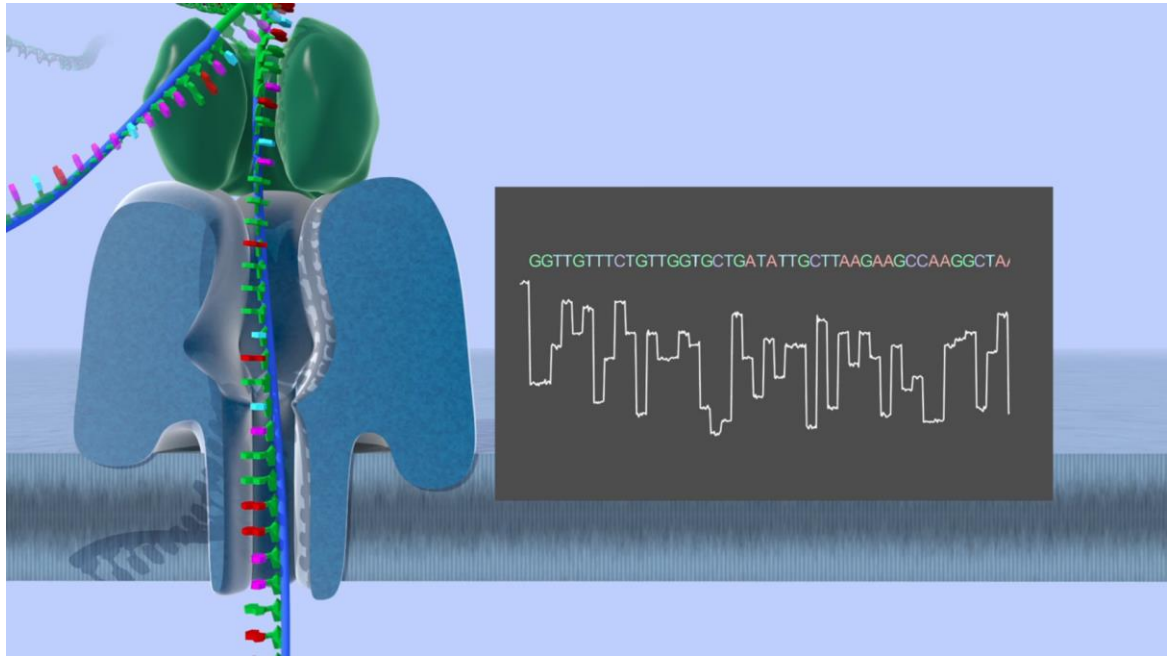
* WGS = whole genome sequencing



Genomics

/nanoporetech/bonito

- Long reads sequencing based on DL network (PyTorch)
- 1D signal processing



https://nanoporetech.com/resource-centre/how-nanopore-sequencing-works-animation

Genomics

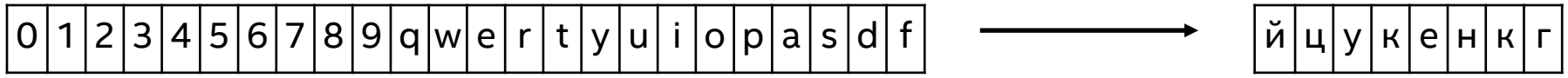
MinION

GridION

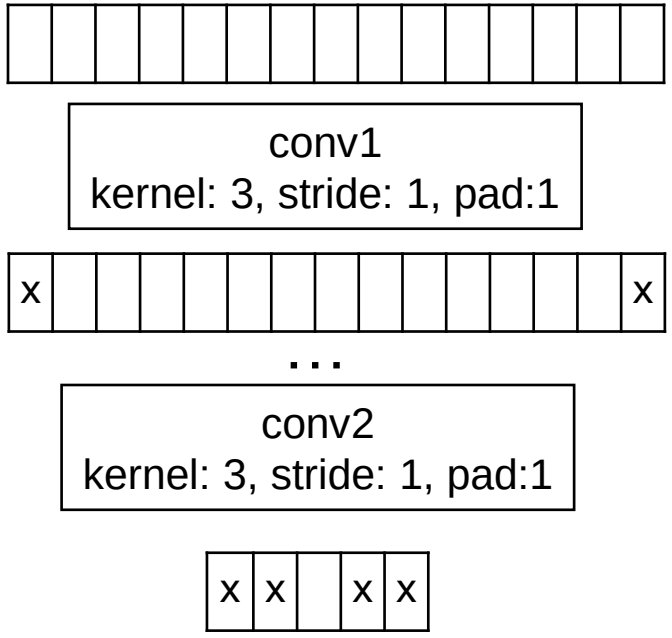
PromethION

Genomics

1. Full read processing – accurate but not scalable



2. Chunks processing



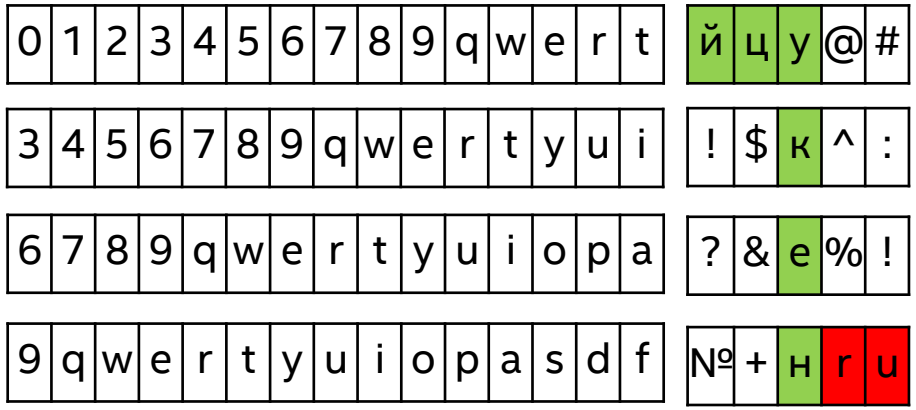
Chunk size – any

overlap – estimate by “clean” area

(depends on the topology, not chunk size)

Chunk size: 15

Overlap: 12



Bonito example: overlap: 2988

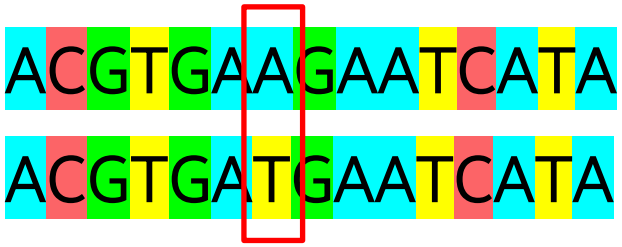
16154 input length is

- 22 chunks of 3600 numbers
- 5 chunks of 6000 numbers

Variant calling

SNPs (snips) - single nucleotide polymorphisms

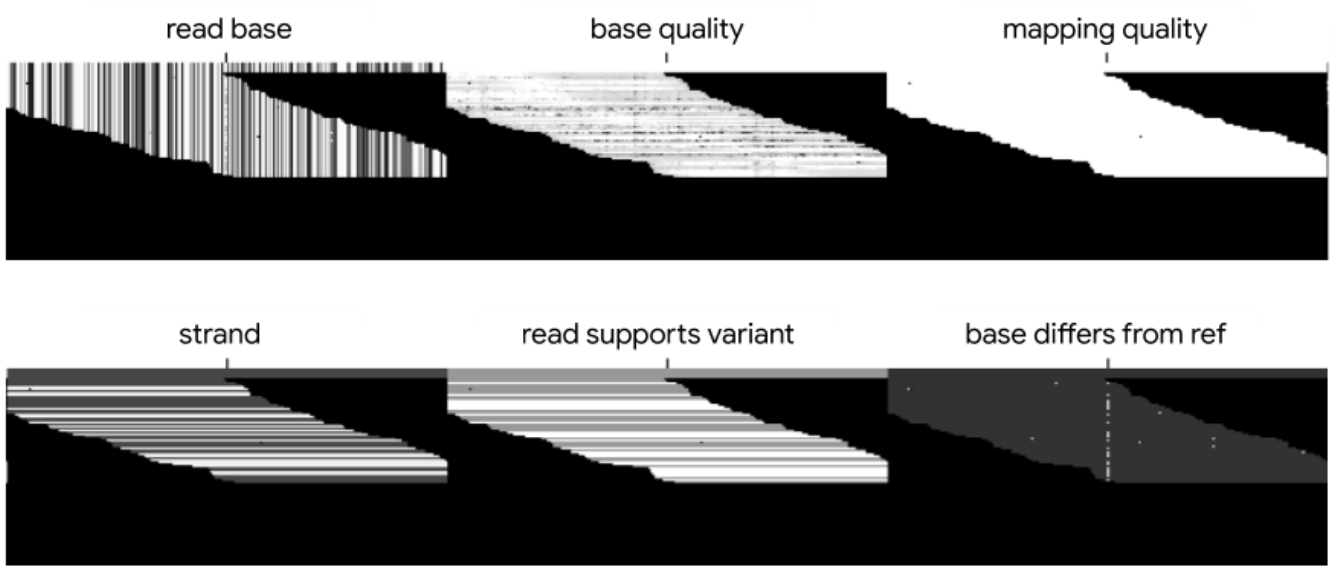
INDEL – insertion or deletion of nucleotides



- Genetic disorder
  - Monogenic (melanoma, phenylketonuria, mucopolysaccharidoses)
  - Polygenic (asthma, coronary heart disease, and diabetes)
- Only 1.5% of genes are “active” (exome)
- *de novo* sequencing: genome functioning rather than structure (i.e. viruses)

Google DeepVariant

- Inception v3 based model
- Docker as primary API



Input: 1x100x221x6

Output classification:

0 - homozygous reference allele (no change)

1 - heterozygous

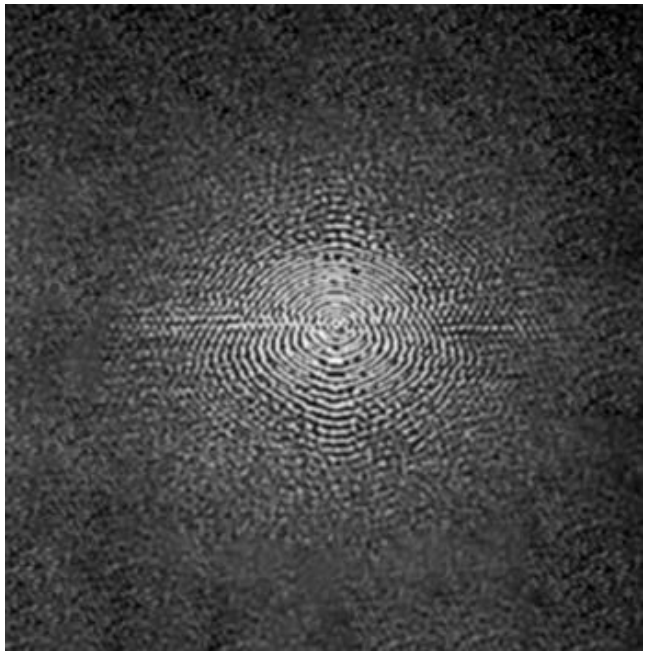
2 - homozygous alternative allele

External Solutions

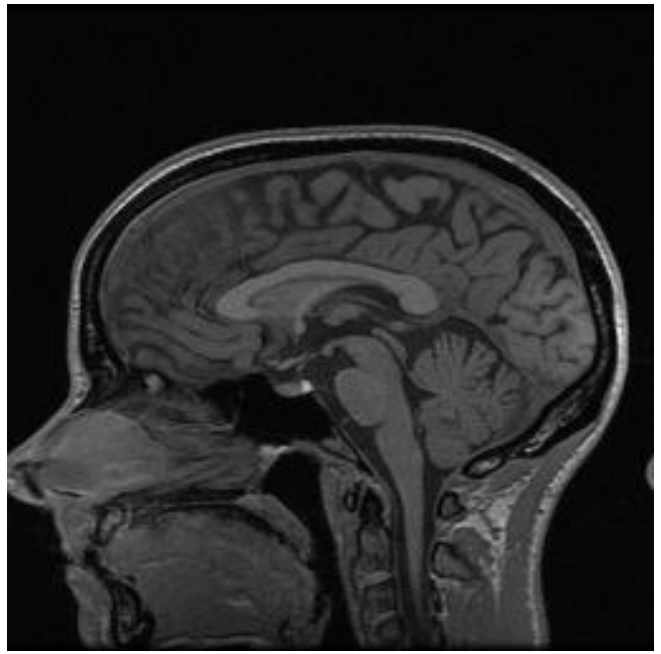
The following pipelines are not created or maintained by the Genomics team in Google Health. Please contact the relevant teams if you have any questions or concerns.

Name	Description
Running DeepVariant on Google Cloud Platform	Docker-based pipelines optimized for cost and speed. Code can be found here.
DeepVariant-on-spark from ATGENOMIX	A germline short variant calling pipeline that runs DeepVariant on Apache Spark at scale with support for multi-GPU clusters (e.g. NVIDIA DGX-1).
NVIDIA Clara Parabricks	An accelerated DeepVariant pipeline with multi-GPU support that runs our WGS pipeline in just 40 minutes, at a cost of $2-$3 per sample. This provides a 7.5x speedup over a 64-core CPU-only machine at lower cost.
DNAnexus DeepVariant App	Offers parallelized execution with a GUI interface (requires platform account).
Nextflow Pipeline	Offers parallel processing of multiple BAMs and Docker support.
DNAstack Pipeline	Cost-optimized DeepVariant pipeline (requires platform account).

Compressed sense



raw signal (k-space)



visualization

Is that possible to sample less MRI / CT data?

- Faster to scan, healthier (physical/mental), lighter storage

Compressed sense

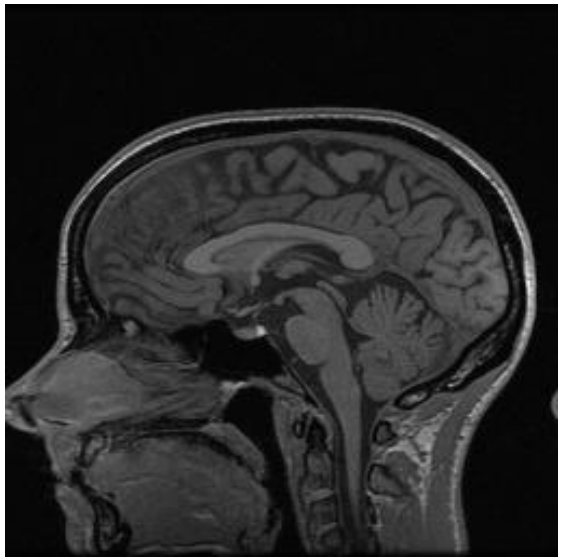
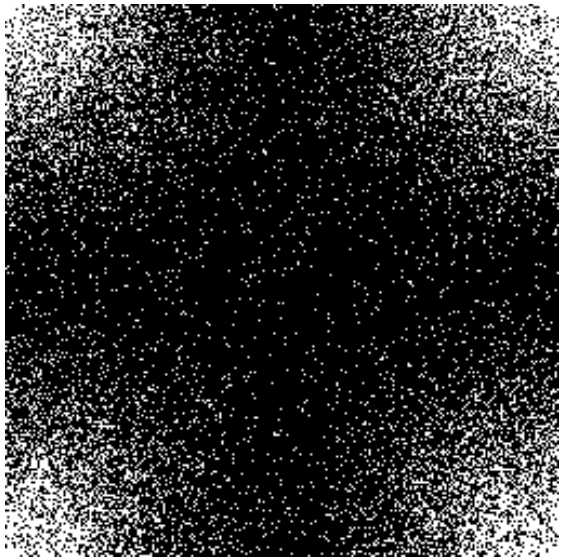
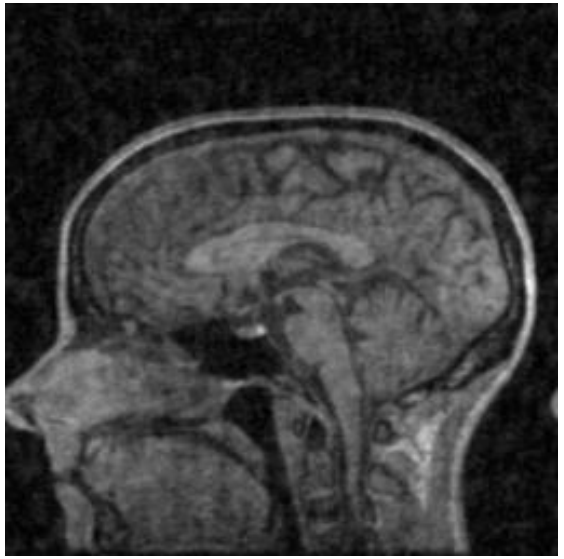


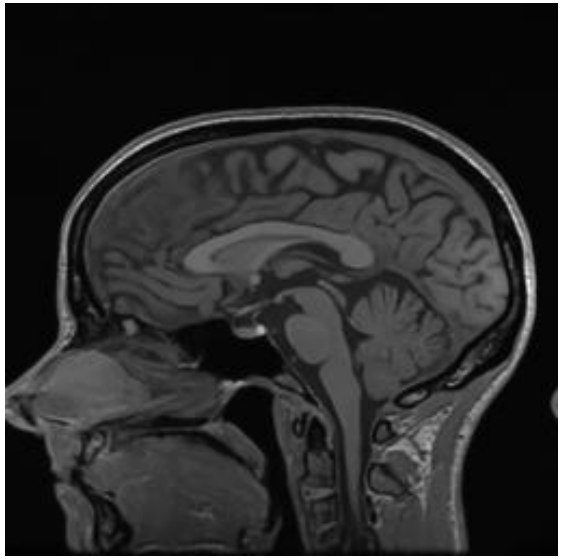
image of origin slice
(from kspace)



sampling mask
0.82 of zeros
(applied to kspace)



slice + mask as is
PSNR: 21.5

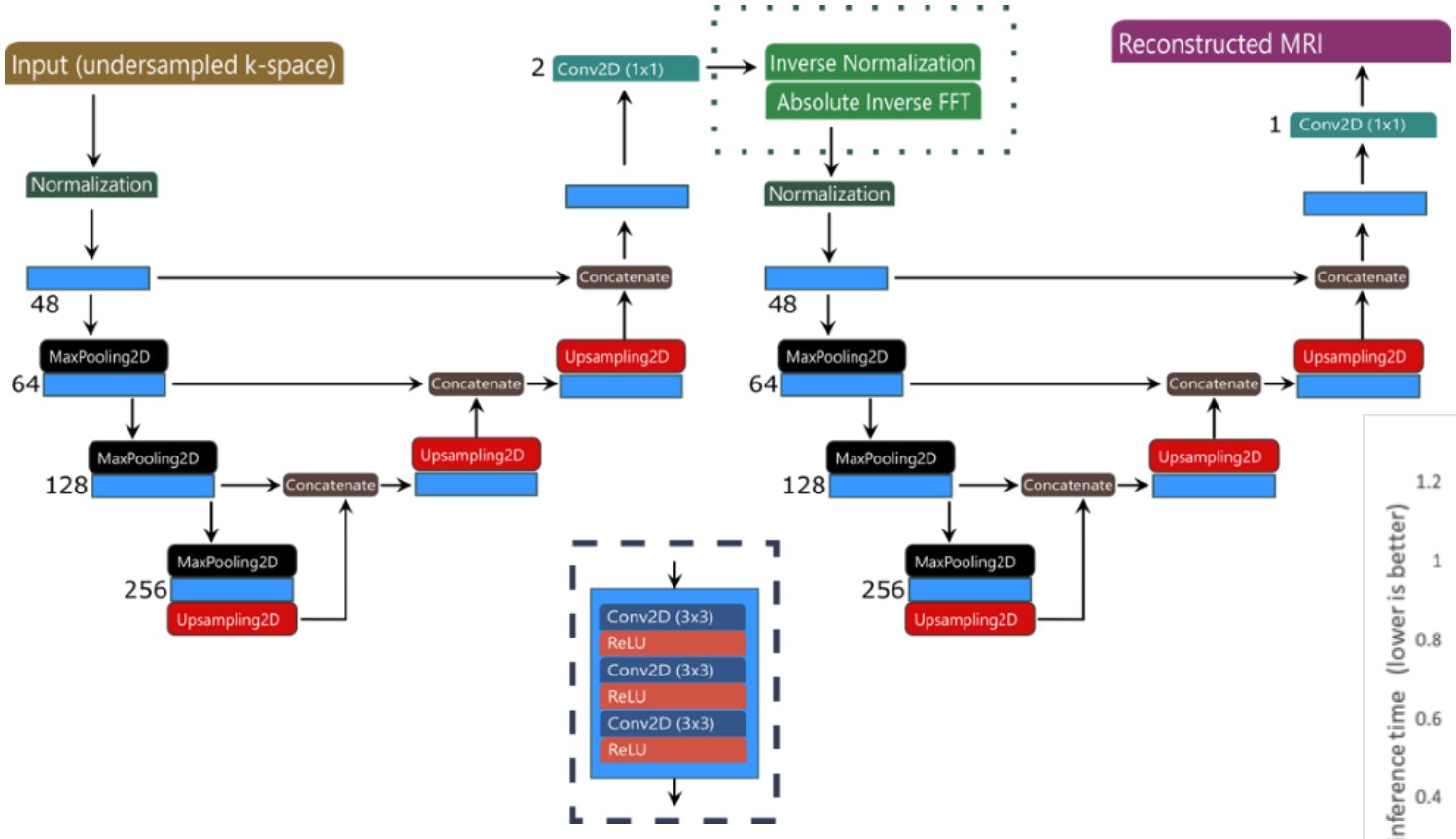


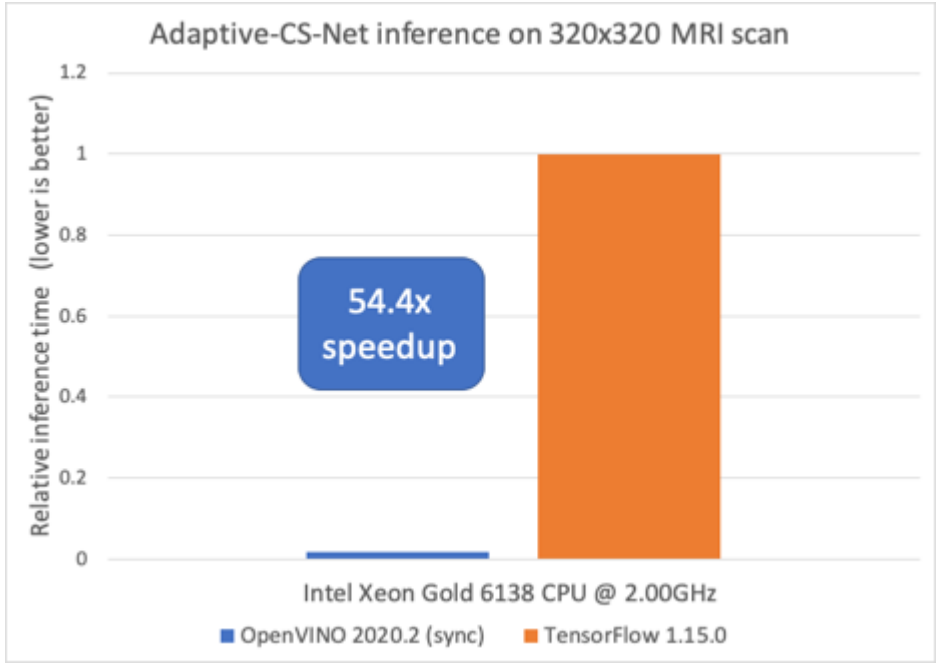
Reconstructed via network
PSNR: 34.8

project (MIT): https://github.com/rmsouza01/Hybrid-CS-Model-MRI/

dataset (Creative Commons Attribution-NoDerivatives 4.0 International Public License):
https://sites.google.com/view/calgary-campinas-dataset/home

Compressed sense

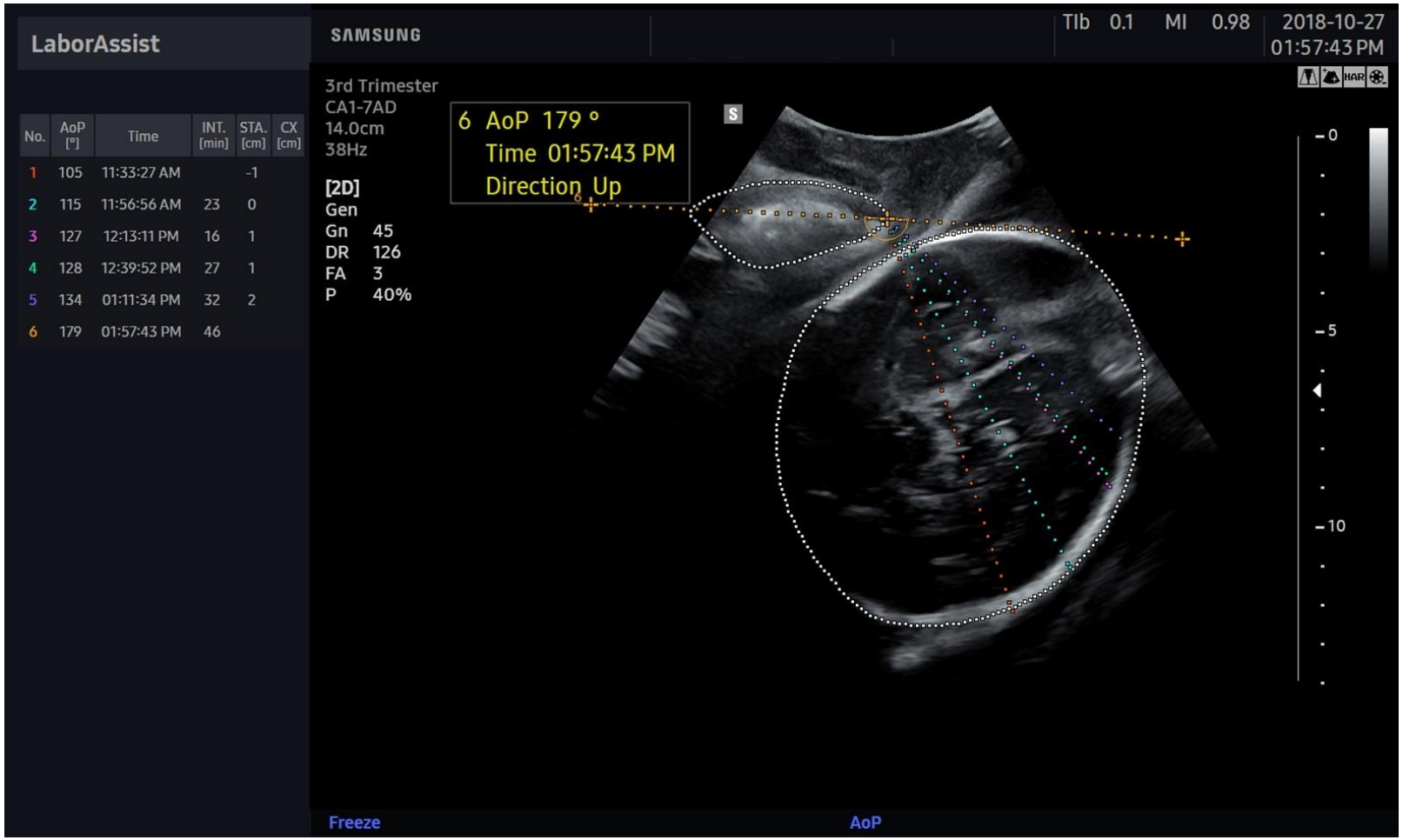




White paper: [Philips Healthcare MRI Image Reconstruction]

Samsung: Fetal Ultrasound Workflow

- Reduce manual operations
- Powered by Intel Core i3



White paper: https://www.intel.com/content/www/us/en/healthcare-it/samsung-medical-imaging-brief.html

intel