mask Intersection-over-Union is blind to the two error sources that decide whether that curve is correct. The first is centreline localisation: a segmentation head emits a probability band several pixels wide at every depth row, and reducing it to a single position by integer per-column argmax snaps the recovered curve to the pixel lattice and caps its accuracy at half a pixel regardless of mask quality. The second is depth registration: the pixel rows of a scan carry depth only through a header scale, and an affine pixel-to-depth map fitted from the header anchors propagates any header mis-read into a depth-dependent value error before any curve interpolation runs. This whitepaper presents the validation framework we built for VeerNet, our encoder-decoder segmentation network for raster well-log digitisation, in which centreline accuracy and depth alignment, not mask IoU, are the acceptance criteria. We define a parabolic three-point sub-pixel refinement of the per-column argmax that recovers fractional position between grid lines, the two-anchor affine depth-registration map and its residual structure, and an end-to-end protocol that grades the reconstructed curve against ground-truth values resampled to a common axis at 300 interpolated depth points using per-curve R-squared, mean absolute error and mean squared error. On that protocol VeerNet reaches a peak coefficient of determination of 0.9891, a lowest mean absolute error of 0.0132 and a lowest mean squared error of 0.0004, even though the same checkpoints score a peak mask IoU of only 0.51. The framework ships three interactive exhibits, a sub-pixel centreline refiner, a depth-registration ladder, and an end-to-end residual scanner, that let a reviewer move each error source independently and read its consequence on the deliverable. We close with the methods needed to reproduce the protocol, its limitations, and the acceptance rule we now apply to every digitised curve.

Tarry Singh
Founder & CEO

Narendra Patwardhan
Research Collaborator

Tannistha Maiti
Senior AI Researcher

Quamer Nasim
ML Research Engineer

VEERNET / RASTER-LOG-DIGITISATION / SUB-PIXEL / CENTRELINE / DEPTH-REGISTRATION / VALIDATION

CONTENTS

> "
>
> A mask can overlap the ground truth poorly and still carry the curve to the right place, or overlap it well and put the value at the wrong depth. Neither failure shows up in IoU. Both show up the moment you grade the value at its depth.
>
> "

I

THE WRONG ACCEPTANCE ARTEFACT

A digitiser is accepted on a mask, delivered as a curve

When a raster-log digitiser is reviewed, the artefact on the table is almost always the segmentation mask, and the number quoted next to it is almost always Intersection-over-Union. We built VeerNet, our encoder-decoder segmentation network for raster well-log digitisation, and for a while we reported it the same way, because that is the habit a segmentation network inherits from the literature it descends from [1]. The habit is convenient and it is wrong for this product class. A petrophysics package never opens the mask. It opens a curve: a value at a measured depth, a column in a Log ASCII Standard file. The acceptance question is not whether the predicted pixels overlap the true pixels. It is whether the value the digitiser reports at a given depth is the value that was actually logged there.

Those two questions come apart, and they come apart in a way that matters. A mask can score a high overlap and still place the curve a few feet too deep, because the rows were mapped to depth from a header tick that was read off by a handful of pixels. A mask can score a poor overlap and still carry the curve to the right value, because the band is wide but its mass sits in the right place. The overlap metric is blind to both. What it measures is the geometric agreement of two sets of pixels, and the deliverable is neither a set of pixels nor a geometry. It is a function from depth to value, and the errors that corrupt that function live in two specific places.

This whitepaper is about those two places and the framework we built to grade them. The first is centreline localisation: turning a probability band several pixels wide into one position per depth row, at sub-pixel precision, rather than snapping it to the pixel grid. The second is depth registration: turning a pixel row into a true measured depth through a map fitted from the scan header, and bounding the error that map can introduce. We treat the accuracy of those two operations as the real acceptance criteria, we validate them end to end against ground truth, and we report the deliverable-space numbers that result. On our held-out validation set, evaluated at 300 interpolated depth points, the reconstructed

curves reach a peak coefficient of determination of **0.9891**, a lowest mean absolute error of **0.0132** and a lowest mean squared error of **0.0004**, while the same checkpoints score a peak mask IoU of only **0.51**. That gap is the whole reason a mask metric cannot be the acceptance test.

II

WHY NOW

Legacy archives are large, and a wrong depth is worse than a missing one

The market context for getting this right is the size and the role of the archives being digitised. An onshore operator that has been drilling for decades sits on tens of thousands of paper and raster logs, and the value of digitising them is not the image; it is the curve fed into petrophysical workflows that were built to consume LAS [7]. A curve that is off by a constant offset in value, or stretched in depth because the header was mis-registered, is not a partial win. It is a liability, because a petrophysicist who trusts a digitised porosity curve at the wrong depth will tie it to the wrong formation top, and the error propagates silently into every interpretation that leans on it. A missing curve announces itself. A confidently wrong one does not.

That asymmetry is why a mask-overlap acceptance test is dangerous rather than merely imprecise. Overlap is a measure of how much of the curve was found, and finding the curve is the easy part. The hard part, the part that decides whether the digitised archive is trustworthy, is placing each value at the right magnitude and the right depth. A validation framework that does not separate and grade those two placements will pass models that overlap well and mislocate quietly, and it will fail models that overlap poorly and reconstruct cleanly. We have seen both, on our own checkpoints, which is what forced us to build the framework this document describes rather than keep tuning toward a mask number.

III

THE CENTRELINE PROBLEM

A band is not a position, and the integer pick wastes the band

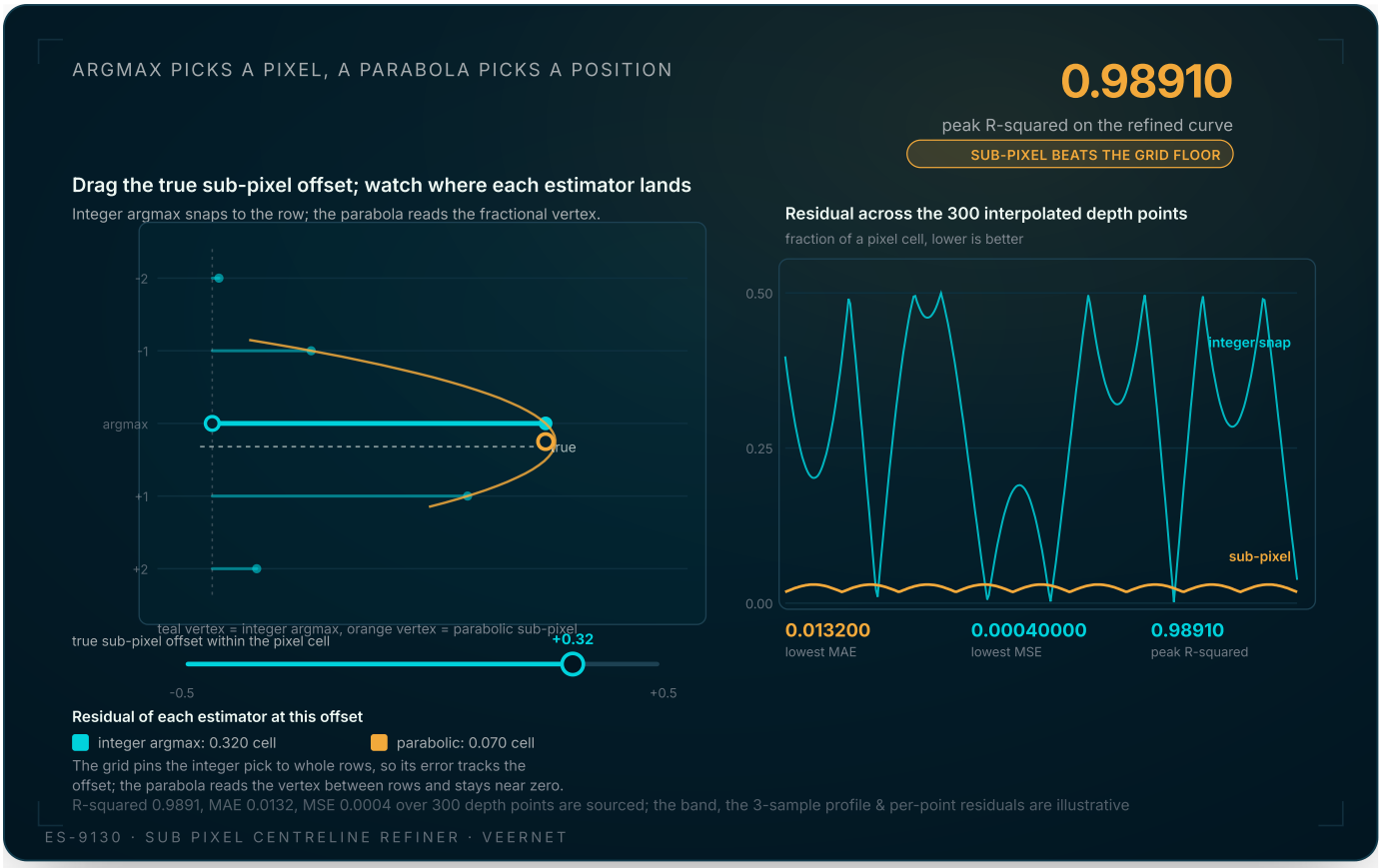
A semantic-segmentation network does not emit a centreline. At every depth row it emits a probability profile across the scan columns, and where a curve trace is present that profile is a soft band a few pixels wide, peaked near the true position and decaying to either side. The deliverable needs one number per row: the horizontal position of the curve at that depth, which becomes, after the depth map, the value at that depth. So the band has to be reduced to a position, and the reduction is where most of the recoverable accuracy is won or lost.

The reflex reduction is the per-column argmax: at each row take the column with the highest probability and keep its integer index. It is one line of code and it is correct to the nearest pixel. That last clause is the problem. Snapping to the nearest pixel quantises the recovered curve onto the integer lattice, and quantisation puts a hard floor under the error that no improvement in the mask can lift. If the true curve sits a third of a pixel below a grid line, the integer argmax will return the grid line, and the residual is that third of a pixel, deterministically, on a clean band and a noisy one alike. Across a log whose curve drifts smoothly through pixel cells with depth, this snapping produces a sawtooth residual locked to the grid, swinging between zero and half a pixel as the curve crosses cell boundaries. The mask can be flawless and the sawtooth is still there.

The fix is older than deep learning and comes from peak interpolation. The discrete band samples around the argmax are samples of a smooth underlying profile, and a smooth peak is locally well approximated by a parabola. So we fit a parabola through the argmax sample and its two immediate neighbours and read the vertex, which lands between the grid lines [2]. The vertex offset has a closed form: with the argmax probability and its lower and upper neighbours, the sub-pixel correction is half the difference of the neighbours divided by the second difference, and adding that correction to the integer index gives a fractional position. This is the same estimator that recovers sub-sample peak locations in spectral analysis, and it carries the same caveat, a known bias when the peak is sharply curved,

which we return to in the limitations [2]. The broader framing, that sub-pixel localisation is an estimation problem solved by fitting a local model rather than reading a discrete maximum, is the one the registration literature has used for decades [3] [4].

The first exhibit lets you move the true sub-pixel offset of a curve within its pixel cell and watch the two estimators respond. The integer argmax stays pinned to the grid line while the parabolic vertex tracks the true position between the lines, and the right-hand panel reads the consequence across the 300 interpolated validation depth points: the integer residual is a sawtooth locked to the lattice, the sub-pixel residual holds a small floor.



A segmentation head does not hand back a curve; at every depth row it hands back a probability band a few pixels wide, and the recovered position is whatever reduction you run on it. The blunt reduction is the per-column argmax: take the brightest pixel and keep its integer row index. That snaps the curve to the pixel lattice and caps its accuracy at half a pixel no matter how clean the band is. The refinement fits a parabola through the argmax sample and its two neighbours and reads the vertex, recovering a fractional position between the grid lines. Drag the slider to set the true sub-pixel offset of the curve inside its pixel cell. The left panel shows the five-row neighbourhood, the discrete band samples (teal), the fitted parabola (orange), the integer argmax pick (teal vertex) and the parabolic vertex (orange), against the true position (pale dashed tick). The read-out under the slider gives each estimator's residual at that offset; the integer error tracks the offset because the grid pins it to whole rows, while the parabolic error stays near zero. The right panel plots both residuals across the 300 interpolated validation depth points: the integer snap leaves a sawtooth locked to the lattice, the sub-pixel estimate holds a small floor. That floor is why the reconstructed curve clears a peak R-squared of 0.9891, a lowest mean absolute error of 0.0132 and a lowest mean squared error of 0.0004 against the ground-truth values, rather than stalling at the grid-quantisation limit. The four headline figures and the 300-point count are sourced from the validation notebooks; the mask band, the three-sample profile and the per-point residual geometry are illustrative.

The argument the exhibit makes is the first half of the framework. The integer pick is not a worse implementation of the same idea as the parabola; it is a different operation that throws away the sub-pixel information the band already contains. The band, by being several pixels wide and smoothly peaked, encodes the fractional position in the relative heights of its samples, and the argmax discards that by keeping only which sample is largest. The parabola reads it back. On its own that is the difference between a curve that

clears R-squared 0.9891 and one that stalls at the grid-quantisation limit, because once the residual is dominated by a fixed half-pixel sawtooth, no amount of training moves the deliverable number.

IV

THE DEPTH PROBLEM

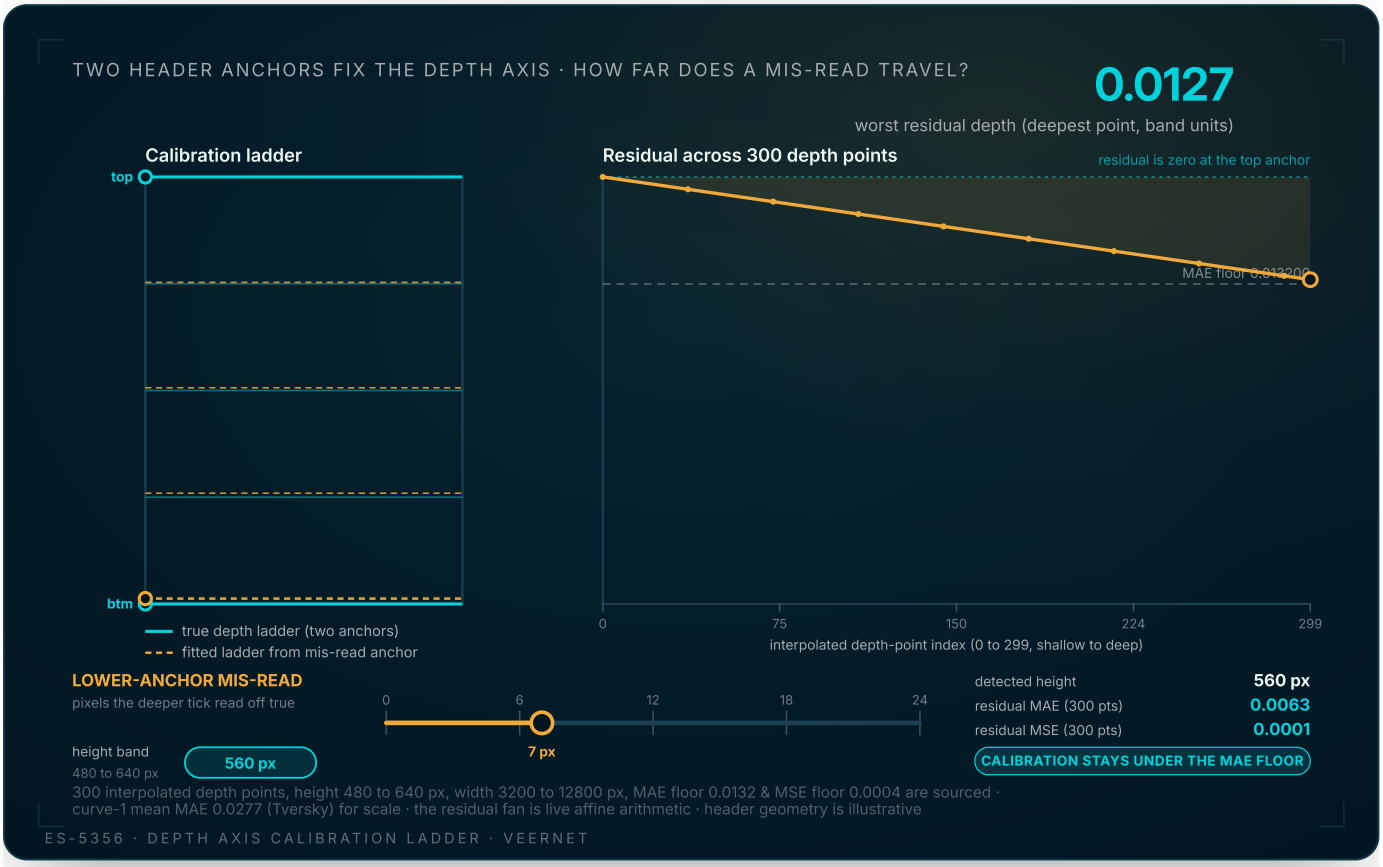
A pixel row is not a depth until the header says so

The second error source is independent of the first and at least as consequential. A sub-pixel-perfect centreline gives the horizontal position of the curve at a row, but a row is a position in image space, not a depth. The depth lives in the scan header, where two printed depth ticks anchor the vertical axis, and recovering depth means fitting the affine map from pixel row to measured depth from those two anchors. The slope of that map is the header depth interval divided by the detected image height in pixels, and the intercept ties the top anchor to its depth. Once fitted, the map converts every detected centreline row into a true depth, and the value-at-depth pairs become the curve that ships.

The danger in this step is that the whole map rests on two detections, and an affine map is exact in those two points and only those two. If the deeper header tick is detected a few pixels off true, the fitted slope is wrong, and the resulting depth error is zero at the top anchor and grows linearly toward the bottom of the log. A curve that is perfectly localised in image space is then reported at progressively wrong depths as you go deeper, and because the error is smooth and monotone it looks plausible. It will not announce itself as a glitch. It will quietly stretch or compress the curve against depth, and tie features to the wrong depths at exactly the deep intervals where a misplacement is most expensive to a petrophysicist.

This is why we evaluate depth registration as its own residual, before any curve interpolation runs, across the same 300 interpolated depth points the end-to-end score uses. The residual structure is diagnostic on its own: a fan opening from zero at the top anchor to its worst value at the deepest point is the signature of a slope error from an anchor mis-read, and the magnitude of that fan tells you whether a header slip would dominate the final CSV or sit beneath the model's own value error. The second exhibit makes that map and its residual fan manipulable. Move the lower-anchor mis-read and the

fitted ladder tilts away from the true ladder, and the residual fan opens, with the read-outs converting the fan into a depth-residual MAE and MSE you can compare directly against the model's own lowest reported MAE of 0.0132 and MSE of 0.0004.



A scanned log carries depth only in its header scale, where two printed depth ticks anchor the vertical axis; calibration is the affine map from a detected pixel row to a true measured depth, fitted from those two anchors with a depth-per-pixel slope equal to the header depth interval over the detected image height. Drag the lower-anchor mis-read lever to detect the deeper header tick a few pixels off true. The fitted ladder (orange) tilts away from the true ladder (teal), and on the right the residual depth error fans open from zero at the top anchor to its worst value at the deepest of the 300 interpolated validation depth points, before any curve interpolation runs. The read-outs convert that fan into a residual MAE and MSE across the 300 points and compare them against the model's own lowest reported MAE of 0.0132 and MSE of 0.0004, so you can see exactly when a header-read slip would dominate the final CSV. The 300 interpolated depth points, the 480 to 640 px height band, the 3200 to 12800 px width range, the MAE and MSE floors, and the curve-1 mean MAE of 0.0277 under Tversky loss are sourced from the engagement archive; the residual fan is live affine arithmetic, and the header tick geometry is an illustrative schematic.

Reading the two exhibits together is the point of separating the errors. A reviewer can hold the centreline perfect and open the depth fan, or hold the depth map perfect and move the sub-pixel offset, and see which source is biting on a given log. A combined mask metric collapses both into a single overlap number that moves for the wrong reasons and stays put

for the wrong reasons, and gives a reviewer no handle on which placement to fix. The framework's discipline is to keep the two error sources on separate axes until the very last step, where they are combined deliberately into the deliverable score.

V

END TO END

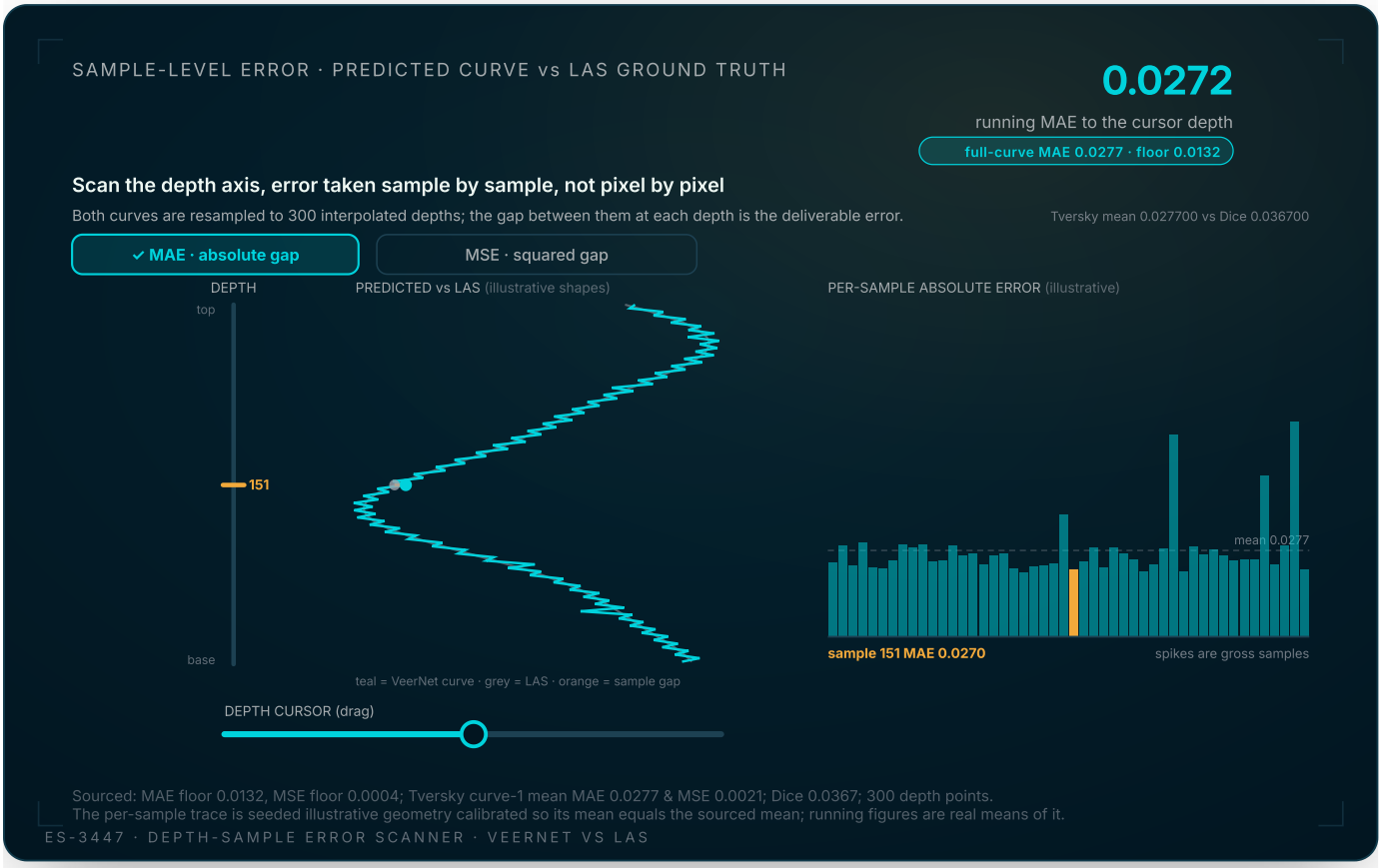
The acceptance number is the curve against ground truth, at depth

The final step composes the two operations and grades the result against ground truth, which is the only number an acceptance review should sign. The predicted curve, sub-pixel centreline through the affine depth map, and the ground-truth curve from the native LAS are both resampled onto a common depth axis, because they do not naturally share one: the prediction is sampled at whatever rows the scan exposed, the ground truth at whatever depths the original tool logged. We bring them onto a shared axis of 300 interpolated depth points using spline resampling [5], and only then compute per-curve R-squared, mean absolute error and mean squared error between the two [6]. Comparing at the same depths is not a detail; comparing two curves at different depths produces a meaningless error, and a framework that skips the resampling step is grading noise.

On that protocol the numbers are the ones quoted throughout: a peak coefficient of determination of 0.9891, a lowest mean absolute error of 0.0132 and a lowest mean squared error of 0.0004, against a peak mask IoU of 0.51 on the same checkpoints. We report all three deliverable metrics rather than one because they fail in different directions. R-squared rewards getting the shape of the curve right, which is what a petrophysicist reads first. MAE reports the typical magnitude of the per-sample error in the curve's own units, the number to quote when asked how far off a single value usually is. MSE squares the error, so it surfaces the rare gross mistake that MAE would average away, the spike at one bad depth that a shape-aware metric and an average-aware metric would both let through. A curve that scores well on all three has its shape, its typical value and its worst value all under control.

The third exhibit is the end-to-end residual scanner: it sweeps the per-sample residual of the reconstructed curve against ground truth across the depth axis and lets you toggle between the MAE and the MSE reading, so the place where a single bad sample spikes the

squared error while barely moving the absolute error is visible directly. This is the exhibit that stands in for the acceptance review itself, the residual a reviewer walks before signing.



The deliverable is graded sample by sample, not pixel by pixel. VeerNet's predicted curve and the LAS ground truth are resampled onto the same axis at 300 interpolated depths, and the gap between the two one-dimensional functions at each depth is the error that survives to the customer. Drag the depth cursor: the running mean accumulates from the top of the curve to the cursor, settling toward the sourced full-curve figures (lowest MAE 0.0132, lowest MSE 0.0004; Tversky curve-1 mean MAE 0.0277, mean MSE 0.0021). Toggle MAE versus MSE to see why both are reported: MSE squares the gap, so the rare gross sample that MAE averages away spikes the squared-error bar. The per-sample trace is seeded illustrative geometry calibrated so its mean equals the sourced mean; all printed figures are either sourced constants or true running means of the trace.

THE REPORTING CONVENTION WE SETTLED ON

Every digitised curve is reported with three deliverable-space numbers against ground truth at the 300 interpolated depth points, R-squared, MAE and MSE, plus the two diagnostic residuals that produced them, the centreline residual and the depth-registration residual. The mask IoU is recorded for training diagnostics and never quoted as an acceptance number.

VI

RESULTS

What the framework reads on our checkpoints

The headline results are the three deliverable metrics, and the contrast that motivates the whole document. The reconstructed curves reach a peak R-squared of 0.9891, a lowest MAE of 0.0132 and a lowest MSE of 0.0004, while the masks those curves were reduced from score a peak IoU of 0.51. Read as a mask, the model looks unfinished. Read as a curve at depth, it clears the bar a downstream petrophysical workflow needs. The framework does not improve the model; it measures the artefact the model actually produces, and that artefact was always better than the mask metric let on.

0.9891
peak R-squared on the reconstructed curve

0.0132
lowest mean absolute error against ground truth

0.0004
lowest mean squared error against ground truth

300
interpolated depth points in the evaluation

The two diagnostic residuals explain how those headline numbers are reached and where they are at risk. The centreline residual is what the sub-pixel refinement buys: with the parabolic vertex the per-row residual collapses toward a small floor instead of the half-pixel sawtooth the integer argmax leaves locked to the grid. The depth-registration residual is the exposure that has to be controlled to keep the headline numbers honest: it is zero when the header anchors are read true and fans open with any anchor slip, so a deep log with a mis-read tick can have a clean centreline and still report wrong values at depth. Keeping both diagnostic residuals beneath the model's own value-error floor is the condition under which the headline R-squared of 0.9891 is trustworthy rather than coincidental.

VII

METHODS

The protocol in the detail needed to reproduce it

The validation framework is a fixed sequence applied to every checkpoint, and the order matters because the steps are not commutative. We state it in full so it can be reproduced.

The input is a held-out scan and its native LAS ground truth. The network produces, per depth row, a probability profile across scan columns [1]. For each row we take the per-column argmax to locate the integer position, then read the argmax probability and its two column neighbours and apply the parabolic vertex correction to recover the sub-pixel position [2]. The correction is the closed-form three-point estimator: half the difference of the neighbour probabilities divided by their second difference, added to the integer index. This produces a sub-pixel centreline, one fractional column position per depth row.

It is worth being explicit about why the three-point estimator is the right amount of model and not more. We deliberately fit a parabola rather than a higher-order polynomial through a wider window, because the parabola has exactly three free parameters and the three samples around the argmax determine them uniquely, with no fitting and no regularisation choice to defend. The vertex is a closed form in two differences of the sample heights: the numerator is the difference between the lower and upper neighbour, which says which side of the argmax the true peak leans toward, and the denominator is the second difference, which measures how sharply the band is curving and therefore how far a given height imbalance moves the vertex. A flat-topped band has a small second difference and the correction is large and uncertain; a sharply peaked band has a large second difference and the correction is small and stable. Reading the estimator this way also tells you when to distrust it, which is precisely when the denominator is near zero, and we clamp the correction in that degenerate case rather than let a near-flat triple emit a wild fractional position.

We then fit the depth-registration map. The two header depth anchors are detected and the affine map from pixel row to measured depth is fitted from them, with the depth-per-pixel slope equal to the header depth interval over the detected image height and the intercept tying the top anchor to its depth [3]. Applying this map to the centreline rows converts the per-row positions into value-at-depth pairs, which is the predicted curve in deliverable form.

The predicted curve and the ground-truth LAS curve are then resampled onto a common depth axis of 300 interpolated depth points by spline interpolation [5], so that both are defined at identical depths. Only after the resampling do we compute the metrics: per-curve R-squared, mean absolute error and mean squared error between the two resampled curves [6]. Alongside the end-to-end metrics we record the two diagnostic residuals, the centreline residual against the band-implied true position and the depth-registration residual against the true anchor depths, each across the same 300 points, so that a poor deliverable score can be attributed to its source.

The choice of 300 interpolated depth points is not arbitrary and it is the step most often skipped. Predicted and ground-truth curves are sampled on different depth grids by construction, so before any error can be computed the two have to be brought onto a shared axis, and the resolution of that axis is a deliberate parameter. We fix it at 300 points spanning the logged interval because that is dense enough to resolve the curve features a petrophysicist reads while staying coarse enough that the resampling does not invent structure between the original samples. Resampling with splines means the comparison is between two smooth reconstructions evaluated at identical depths [5], which is the only comparison that has a defined meaning; scoring a prediction against ground truth at whatever rows the scan happened to expose would compute the difference of two curves at different depths, and the resulting number, however small, would be measuring depth mismatch rather than value error. Stating the point count alongside every metric is part of the discipline, because an R-squared at 300 points and an R-squared at a thousand points are not the same quantity, and a data sheet that omits the count omits half the metric.

The two operations are validated independently before they are composed. The centreline estimator is checked by sweeping a known sub-pixel offset and confirming the parabolic vertex recovers it within the estimator's bias, against the integer argmax which cannot. The depth map is checked by perturbing an anchor detection by a known pixel error and

confirming the residual fan matches the predicted linear-in-depth structure. Composing two operations that have each been validated in isolation is what lets us read the end-to-end number as a property of the deliverable rather than an accident of two errors cancelling.

VIII

SENSITIVITY

Which error source dominates, and when

The practical use of separating the two residuals is knowing which one to fix on a given log, because they dominate in different regimes. On a short, cleanly scanned log with sharp header ticks, the depth map is nearly exact and the centreline residual dominates, so the sub-pixel refinement is where the accuracy is won and the integer argmax is the thing to never ship. On a long log, or one whose header ticks are faint or skewed, the depth-registration residual fans open with depth and can overtake even a well-controlled centreline error at the deep end, so anchor detection becomes the binding constraint and a sub-pixel-perfect centreline buys nothing below the point where the depth fan dominates.

This is why a single combined metric is not just imprecise but actively misleading for triage. A mask IoU that drops could be a centreline problem the refinement would fix, or a depth problem the refinement cannot touch, and the IoU gives no way to tell. The two diagnostic residuals do: a flat depth residual with a sawtooth centreline residual says fix the reduction; a clean centreline residual with an opening depth fan says fix the header detection. The framework turns acceptance from a pass-fail on one opaque number into a diagnosis with a named cause, which is the property that lets an engineering team act on a failed log instead of retraining and hoping.

IX

IMPLICATIONS

What changes when the curve at depth is the unit of account

Treating centreline accuracy and depth alignment as the acceptance criteria, rather than mask overlap, changes three things about how a digitiser is built and bought. It changes what gets optimised, because once the deliverable is the curve at depth, effort flows to the reduction and the registration rather than to squeezing the last points of IoU out of a band whose overlap was never the deliverable. It changes how a model is reported, because the data-sheet number becomes the per-curve R-squared, MAE and MSE against ground truth at a stated number of depth points, with the mask IoU demoted to a training diagnostic. And it changes the conversation in procurement, because a buyer who knows the two error sources can ask the two right questions, how is the centreline reduced and how is depth registered, instead of negotiating on an overlap figure that does not predict whether the delivered curves will be usable.

None of this requires a better model. The checkpoints that reach R-squared 0.9891 at the deliverable are the same checkpoints that score IoU 0.51 at the mask. What the framework provides is a measurement that sees the deliverable, separates the two placements that decide whether it is correct, and grades each one against ground truth at the depths a petrophysicist will actually read. The model was always producing a good curve. The mask metric just could not see it, and a poorly registered depth axis could have quietly ruined it.

The acceptance test we run now is short to state and unforgiving to pass: reduce the band to a sub-pixel centreline, register the rows to true depth from the header, resample both curves to a shared depth axis, and grade R-squared, MAE and MSE against ground truth, with the two diagnostic residuals on record to name the cause of any miss. A digitised curve that survives that test is one a petrophysicist can tie to a formation top without second-guessing the depth, and that, not pixel overlap, is the only definition of done we are willing to sign.

KEY TAKEAWAYS

1. Mask IoU is blind to the two errors that decide a digitised curve: where the centreline sits inside the band, and how pixel rows map to depth.
2. Integer per-column argmax snaps the curve to the pixel grid and caps accuracy at half a pixel; a parabolic three-point vertex recovers the fractional position.
3. Depth registration is an affine map from two header anchors; a mis-read anchor produces a depth error that is zero at the top and fans open with depth.
4. Grade the deliverable, not the mask: resample predicted and ground-truth curves to 300 common depth points, then score R-squared, MAE and MSE.
5. The same checkpoints score peak IoU 0.51 yet reach R-squared 0.9891, MAE 0.0132 and MSE 0.0004 once the curve is graded at depth.
6. Keep the centreline residual and the depth residual on separate axes so a failed log gets a named cause, not an opaque drop in one number.

GLOSSARY

Centreline	The single curve position recovered per depth row from the predicted probability band. A band is several pixels wide; the centreline is the one value that ships, so how it is estimated sets the floor on curve accuracy.
Depth registration	The affine map from a detected pixel row to a true measured depth, fitted from the two printed header anchors. Any header mis-read tilts the map and propagates a depth-dependent value error into the final CSV.
Interpolated depth points	The 300 points on a common depth axis to which both predicted and ground-truth curves are resampled before scoring, so the two are compared at the same depths rather than at whatever rows the scan happened to expose.
MAE	Mean absolute error between predicted and ground-truth values, in the curve's own units, at the 300 interpolated depth points. Lowest here: 0.0132.
MSE	Mean squared error between predicted and ground-truth values. Penalises rare gross errors that MAE averages away. Lowest here: 0.0004.
Parabolic refinement	A three-point vertex fit through the argmax sample and its two neighbours that reads a fractional position between grid lines. The same quadratic peak-interpolation estimator used in spectral peak picking, applied per column.
Per-column argmax	The blunt centreline reduction: at each depth row pick the brightest pixel and keep its integer row index. Correct to the nearest pixel, which caps the recovered curve at half a pixel of error no matter how clean the band is.

R-squared	Coefficient of determination between the reconstructed curve and ground truth on the common depth axis. The deliverable-space headline. Peak here: 0.9891.

Limitations

The framework grades the deliverable well, but its components carry assumptions that bound where the numbers hold. The parabolic three-point estimator is biased when the underlying band peak is sharply curved relative to the pixel spacing, the same bias documented for quadratic peak interpolation in spectral analysis [2], so on very narrow bands the recovered sub-pixel position is pulled slightly toward the integer index and the centreline residual we report is an under-estimate of the achievable precision rather than an exact figure. The depth-registration map is affine and fitted from exactly two anchors, which assumes the scan has no non-linear vertical distortion from skew, paper stretch or scanner non-uniformity; where such distortion is present the residual fan is not the simple linear-in-depth structure the diagnostic assumes, and a two-anchor map will mis-state depth in the middle of the log even when both anchors are read true. The end-to-end metrics depend on the spline resampling onto the common depth axis [5], and at the 300-point resolution we use, features finer than the point spacing are not resolved, so a curve with very high-frequency structure could pass the resampled comparison while differing from ground truth between sampled depths. The R-squared of 0.9891, MAE of 0.0132 and MSE of 0.0004 are peak and lowest figures from the validation set, not population means across all logs, and the mask IoU of 0.51 is a peak on the same set; a log outside the distribution the validation set represents may register larger residuals from either error source. Finally, the framework grades a curve against ground-truth LAS, so it is only as trustworthy as that ground truth, and on logs where the native LAS is itself depth-shifted or noisy the residuals it reports will inherit that error.

References

1. Ronneberger, O., Fischer, P., Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. MICCAI. https://arxiv.org/abs/1505.04597

2. Smith, J. O. (2011). *Spectral Audio Signal Processing*, Quadratic Interpolation of Spectral Peaks. W3K Publishing. https://ccrma.stanford.edu/~jos/sasp/Quadratic_Interpolation_Spectral_Peaks.html

3 Thevenaz, P., Ruttimann, U. E., Unser, M. (1998). *A Pyramid Approach to Subpixel Registration Based on Intensity*. IEEE Transactions on Image Processing, 7(1), 27-41. https://bigwww.epfl.ch/publications/thevenaz9801.html

4 Lucchese, L., Mitra, S. K. (2002). *Using saddle points for subpixel feature detection in camera calibration targets*. Asia-Pacific Conference on Circuits and Systems. https://ieeexplore.ieee.org/document/1115289

5 de Boor, C. (1978). *A Practical Guide to Splines*. Applied Mathematical Sciences, Vol. 27. Springer-Verlag. https://link.springer.com/book/10.1007/978-1-4612-6333-3

6 Virtanen, P., et al. (2020). *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, 17, 261-272. https://www.nature.com/articles/s41592-019-0686-2

7 Koroteev, D., Tekic, Z. (2021). *Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future*. Energy and AI. https://www.sciencedirect.com/science/article/pii/S2666546820300033

Get the full whitepaper

This page is the long-form summary. The complete whitepaper adds the closed-form derivation and bias analysis of the parabolic centreline estimator, the full affine depth-registration arithmetic with worked anchor-error propagation, the spline-resampling and metric definitions used at the 300 interpolated depth points, and the per-curve, per-loss reconciliation between the mask metrics and the deliverable metrics on the same held-out examples.

EARTHSCAN

Sub-Pixel Centrelines and Depth Registration: A Validation Framework for Digitised Curves

Published May 2023. Generated 2026-07-22.
https://www.earthscan.io/whitepapers/sub-pixel-centreline-depth-registration-validation-framework