GeoBFDT: End-to-End Detection Transformers for Fracture and Bedding Picking in Carbonate Image Logs

Automated fracture and bedding picking on borehole image logs has been dominated by mask-based detectors that find one geological feature per image and lean on hand-tuned post-processing — a poor fit for fractured carbonates, where sinusoids overlap, intersect, and crowd a single patch. We present GeoBFDT, an end-to-end Detection Transformer that reframes the task as set prediction: a lightweight ResNet-10 backbone trained from scratch feeds a four-layer transformer encoder and four-layer decoder, a fixed bank of learned object queries emerges as candidate detections, and a Hungarian bipartite-matching loss assigns each query to at most one ground-truth sinusoid before any gradient is computed. The model regresses depth, dip and azimuth directly from the image — no masks, no anchors, no non-maximum suppression. Trained on 14 vertical wells (two different microresistivity imaging tools) from a fractured Middle East carbonate play we partnered on, GeoBFDT reaches an F1 of roughly 65% for fractures and 63% for beddings at a 3 cm depth tolerance, rising to roughly 75% and 69% at 5 cm, with dip accuracy near 90% at 3 degrees and azimuth accuracy near 92% at 15 degrees. Ablations isolate the choices that matter: a ResNet-10 backbone outperforms ResNet-18 and ResNet-34 by an order of magnitude on classification error on this small-data regime, sinusoid-aware augmentation moves classification error from 100% to 2.6%, and dynamic-image input beats static. We argue that set prediction — not bigger masks — is the correct inductive bias for multi-fracture carbonate image logs, and that the engineering, not the geoscience, is what makes it production-grade.

Quamer Nasim
ML Research Engineer

Tannistha Maiti
Senior AI Researcher

Tarry Singh
Founder & CEO

GEOBFDT / DETECTION-TRANSFORMER / DETR / SET-PREDICTION / BIPARTITE-MATCHING / HUNGARIAN-ALGORITHM

CONTENTS

A fracture cutting a borehole wall is, geometrically, a plane intersecting a cylinder. Unroll the cylinder into a flat image and that plane becomes a sinusoid whose amplitude encodes dip and whose phase encodes azimuth. A structural geologist picks these sinusoids by eye, one trace at a time, down kilometres of high-resolution image log. It is careful, expert, and unrelentingly manual work, and in a fractured carbonate it is also where the interpretation backlog accumulates: the intervals that matter most for flow are precisely the ones where fractures overlap, cross, and crowd a single image patch.

The machine-learning question is deceptively narrow. How do you teach a model to output *a set* of geological objects — an unordered collection of sinusoids, each carrying its own depth, dip, and azimuth, three in one patch and zero in the next — and to do it without the segmentation masks and hand-tuned post-processing that the first generation of deep-learning detectors imported wholesale from natural-image computer vision? This whitepaper is about the answer we built and validated in our work with a Middle East NOC / carbonate operator we partnered with: **GeoBFDT**, an end-to-end Detection Transformer that treats fracture and bedding picking as set prediction and regresses the geological parameters directly from the image. It is written for the chief geoscientist, the structural geologists and geomodellers who consume the picks, and the R&D geoscience leads deciding whether this class of model belongs in their workflow.

Why one feature per image is the wrong frame

The strongest published baselines on this problem are mask-based instance detectors — Mask R-CNN and its descendants. They work, and on single, well-separated features they work well: prior studies report a mean intersection-over-union around 81% and, in another, precision near 96% with recall near 92% on borehole-image fracture tasks; acoustic-image methods have reported AUC near 98% on synthetic, single-event data. Those are real results. They are also results on a framing that does not survive contact with a fractured carbonate.

Three structural problems recur. First, **a mask is the wrong output**. The thing a geomodeller needs is not a pixel blob; it is three numbers — depth, dip, azimuth — and a mask reintroduces those as a downstream regression you then have to fit, threshold, and clean. Second, **one feature per image does not match the geology**. The mask-detector pipelines were largely demonstrated on isolated features; in a highly fractured interval a single patch routinely carries several sinusoids that cross each other, and a per-instance mask pipeline

either misses the overlaps or fragments them. Third, **the post-processing is a thicket of thresholds**. Anchor scales, IoU cutoffs, and non-maximum suppression all carry tunable knobs with no natural meaning for a sine wave — and NMS, whose entire job is to delete overlapping detections, will happily delete the real, genuinely-overlapping fractures you most want to keep.

> **THE REFRAMING IN ONE SENTENCE**
>
> Stop asking the model to paint a mask of one feature and clean it up afterward; ask it to emit, in a single forward pass, an unordered set of fractures and bedding planes with their depth, dip, and azimuth already regressed — and score the set as a whole.

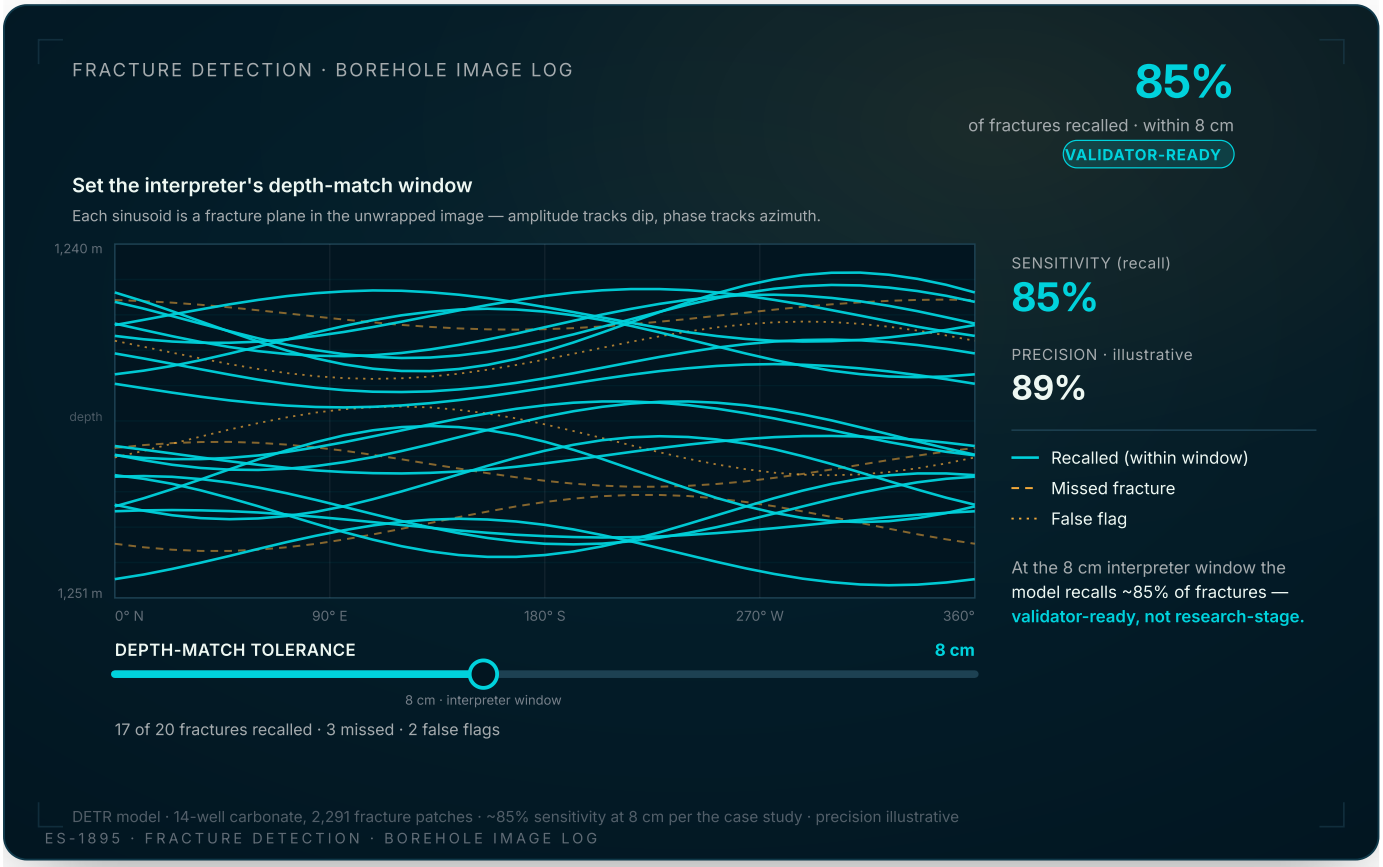
That reframing is exactly what the Detection Transformer formulation provides, and it is why GeoBFDT is built on DETR rather than on a region-proposal network. The shift is not cosmetic. It changes what the loss optimises, it removes the post-processing surface entirely, and — as the ablations below show — it is what lets a deliberately small model generalise on a deliberately small dataset.

The architecture: small backbone, set-prediction head

GeoBFDT is a compact, single-pass computer-vision pipeline. The image-feature stage is a **ResNet-10 convolutional backbone trained from scratch with no pretrained weights** — an unusual choice that the data forced, and one we return to in the ablations. The backbone's feature map, with fixed sinusoidal positional encodings, feeds a **four-layer transformer encoder and a four-layer transformer decoder** with a feed-forward dimension of 1024. The decoder is handed a fixed bank of learned **object queries**; each query attends over the encoded image and emerges as one candidate detection. Two lightweight heads finish the job: a single-linear-layer classification head that labels each query (fracture, bedding, or no-object) and a two-linear-layer regression head that outputs the sinusoid's depth, dip, and azimuth.

Crucially, the geological parameters are regressed *directly*. Depth, dip, and azimuth are each normalised to the unit interval — divided by their physical ranges (100, 90, and 360 respectively) — so that no single parameter dominates the regression gradient, and the

network learns them jointly with the classification. There is no mask, no anchor grid, and no non-maximum suppression anywhere in the pipeline. At inference, the model simply keeps the queries whose object probability clears a **0.5 threshold**; because the training loss guarantees one-to-one assignment, each real fracture is already claimed by exactly one query, so there are no duplicates to suppress.



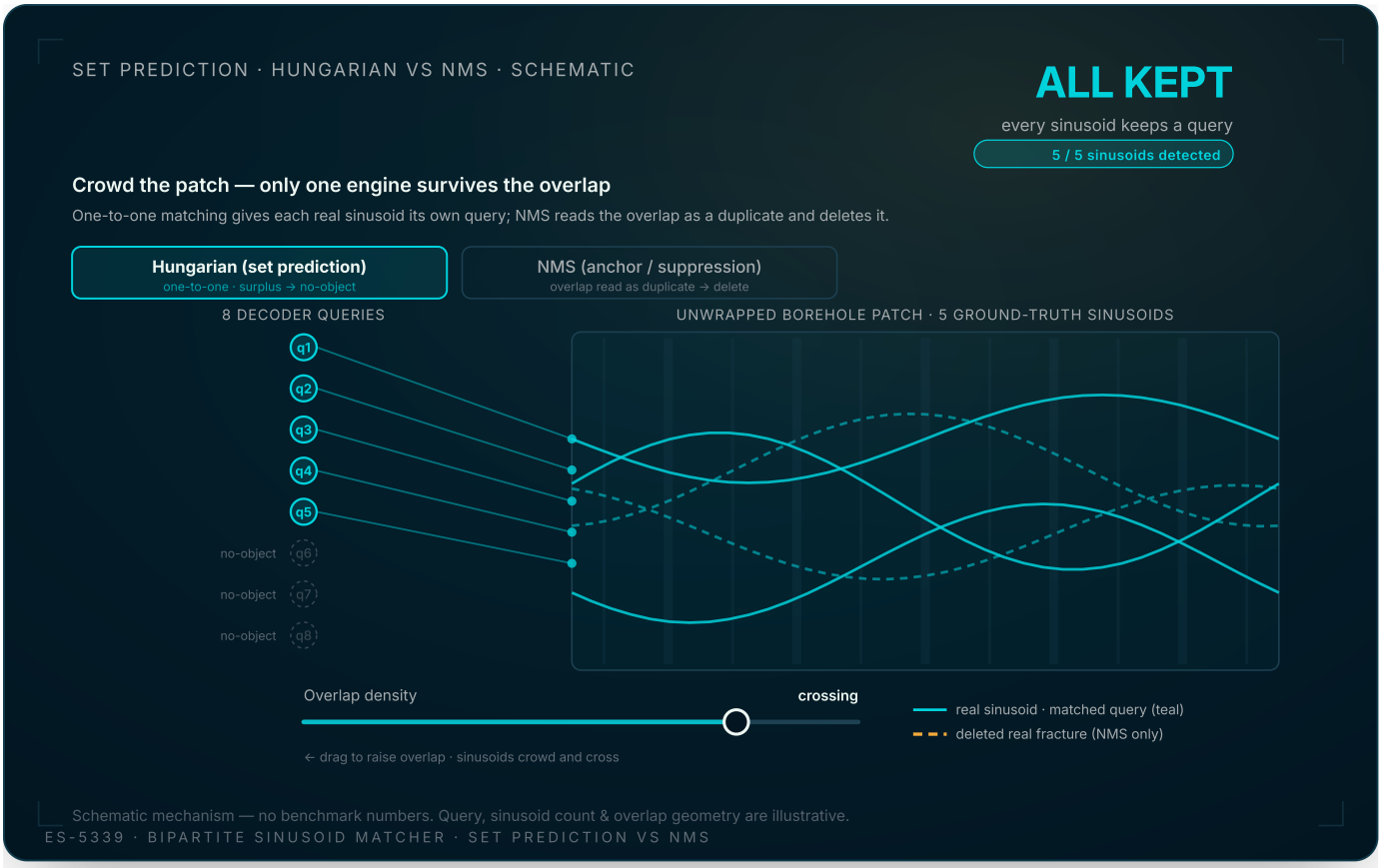
An unwrapped borehole image log: azimuth runs 0–360° across, depth runs down, and every fracture plane intersecting the wellbore traces a sinusoid (amplitude tracks dip, phase tracks azimuth). The DETR model predicts each sinusoid's depth; a pick is a hit only if it falls inside the interpreter's depth-match window. Drag the window — fractures flip recalled (teal) / missed (orange), false flags creep in as it loosens, and sensitivity trades against precision. At the 8 cm interpreter window the model recalls ~85%: validator-ready. Sensitivity, corpus and geometry per the case study; the per-pick errors and live precision are an illustrative detection model around that point.

The live detector above makes the output concrete: on a high-resolution borehole image-log patch, GeoBFDT emits a handful of sinusoids at once, each already carrying its regressed depth, dip, and azimuth — the form a geomodeller can consume directly, not a mask awaiting a second model. The engineering economy is deliberate. A ResNet-10 backbone, a 4+4 encoder/decoder, and two thin heads is a small network by modern standards, and that smallness is a feature on a 14-well dataset, not a compromise.

The loss is where the inductive bias lives

The reframing only pays off because of how GeoBFDT is *scored*. The model always emits the same fixed number of candidate detections regardless of how many fractures a patch actually contains. That immediately raises an assignment problem: if predictions are an unordered set and ground truth is an unordered set, which prediction is graded against which label? There is no canonical order; query number seven has no privileged relationship to the seventh fracture, and there may be no seventh fracture at all.

GeoBFDT solves this by making assignment *part of the loss*. Before any gradient is computed, it builds a cost matrix between every prediction and every ground-truth sinusoid — blending classification confidence with how close the regressed depth, dip, and azimuth land to the target — and runs the **Hungarian algorithm** to find the single lowest-cost one-to-one pairing. Only then is the loss computed on the matched pairs: a **focal loss** on classification and an **L1 loss** on the regressed parameters, weighted with a class-loss weight of 5 against a parameter-loss weight of 1. Focal loss is doing real work here: across a patch the no-object case overwhelmingly dominates, exactly the foreground-background imbalance focal loss was designed for. Unmatched queries are trained, via the classification term alone, to confidently say "no fracture here."



Why set prediction survives overlap where suppression fails. Left: a fixed bank of decoder queries. Right: a schematic unwrapped borehole patch carrying overlapping ground-truth sinusoids (fractures and beddings). Drag the slider to raise overlap density and toggle the engine: Hungarian matching re-solves a one-to-one assignment so every real sinusoid keeps its own query (surplus queries route to 'no-object', shown dimmed) — overlap is the signal, nothing is deleted; flip to NMS and rising overlap makes non-maximum suppression read the tightest-overlapping pair as a duplicate and delete a real fracture (the orange trace). This is a structural mechanism illustration — it sources no benchmark numbers; query count, sinusoid count, and overlap geometry are schematic.

The matcher above is the heart of the model. It is what lets GeoBFDT handle intersecting sinusoids that an NMS-based pipeline would collapse: the one-to-one constraint forbids two predictions from both claiming the same ground-truth fracture, and equally forbids one prediction from absorbing two crossing fractures into a single detection. The geological objects stay separate because the optimisation problem makes them stay separate.

We can write the regression target compactly. A single fracture sinusoid on the unrolled image, before normalisation, takes the form:

FRACTURE SINUSOID ON THE UNROLLED BOREHOLE IMAGE

Formula | LaTeX

Copy LaTeX

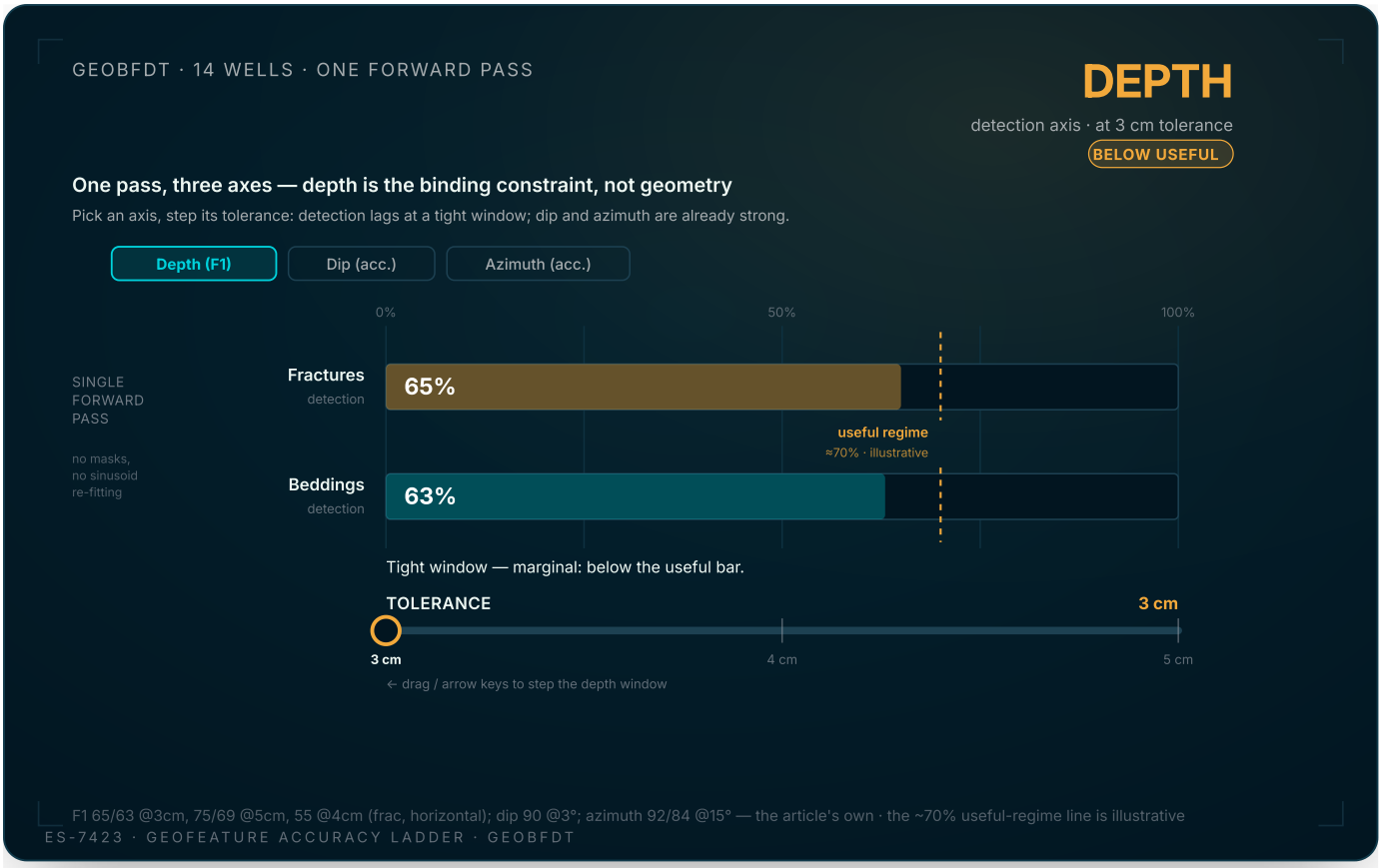
$$y(x) = A\,\sin(\omega x + \varphi) + \text{offset}$$

where the amplitude `A` maps to dip, the phase shift maps to azimuth, and the vertical offset locates the feature in depth. GeoBFDT does not fit this curve with a classical least-squares routine per detection; it regresses the underlying parameters end-to-end, so the same forward pass that decides *whether* a query is a fracture also decides *where, how steep, and which way* it dips.

What the field metrics actually say

Trained on **14 vertical wells** (two different microresistivity imaging tools) from a fractured Middle East carbonate play, GeoBFDT was evaluated against geoscientist-validated picks. The dataset is organised into a fractures-only set drawn from all 14 wells (2,291 non-overlapping images, 1,217 of them carrying fractures) and a combined bedding-and-fracture set from the 11 wells with consistent bedding picks (1,492 images, 1,322 with beddings and 754 with fractures).

The headline detection numbers are deliberately reported against an explicit localisation tolerance, because on this physics the tolerance is not optional. At image resolution, **a single pixel of the digital log format corresponds to about 3 cm of depth**, so a ±3 cm uncertainty is baked into the input before the model does anything — a number 3 cm off a ground-truth pick may be the instrument, not the network. Read against that floor, GeoBFDT reaches an **F1 of roughly 65% for fractures and 63% for beddings at a 3 cm tolerance, rising to roughly 75% and 69% at 5 cm**. On geometry — which is what feeds the structural model — the numbers are stronger and arguably more decision-relevant: **dip accuracy near 90% at a 3-degree threshold (about 80% at 1.5 degrees), and azimuth accuracy near 92% at 15 degrees (about 83% at 7 degrees)** for fractures, with beddings close behind. Aggregated as a single combined-accuracy figure, the model lands near **77% on fractures and 74% on beddings**, with fracture depth, dip and azimuth mean-absolute errors of roughly 1.7 cm, 1.7 degrees, and 9.3 degrees.



GeoBFDT emits the whole (class, depth, dip, azimuth) tuple in one forward pass — but the three axes are not equally hard. Detection along the depth axis is the binding constraint: at a tight 3 cm window fracture F1 is only ~65% (beddings ~63%) and only clears the useful regime for structural work once tolerance loosens to 5 cm (~75% / ~69%); horizontal wells hold ~55% at 4 cm. The geometric axes the interpreter actually fits sinusoids for are already strong at tight tolerance — dip ~90% at 3°, azimuth ~92% (fractures) / ~84% (beddings) at 15°. Pick an axis and step its tolerance: depth is the lever you loosen, dip and azimuth are already past the line. All accuracies and tolerances are the article's own; the dashed ~70% 'useful regime' line is an illustrative reading aid (the article names no exact F1 cutoff).

The ladder above is the honest way to read these results. A single F1 headline is meaningless without its tolerance, and the climb from 3 cm to 5 cm is steep precisely because the 3 cm rung sits right on the instrument's own resolution floor. The point of the chart is not to cherry-pick the most flattering rung; it is to show that the metric is being read *against the physics* rather than in spite of it — and that the geometry channels (dip and azimuth) clear the bar that downstream structural and geomodelling work actually cares about.

A note on comparison to the mask-based baselines: their headline numbers (mIoU ~81%, precision ~96%/recall ~92%) are not directly commensurable with GeoBFDT's F1-at-tolerance, because they measure mask overlap on largely single-feature tasks rather than set-level detection of multiple, overlapping sinusoids with directly regressed geometry.

GeoBFDT trades a few points of nominal overlap score for something the mask pipelines structurally cannot deliver: multiple intersecting features per patch, the geological parameters as first-class outputs, and zero post-processing.

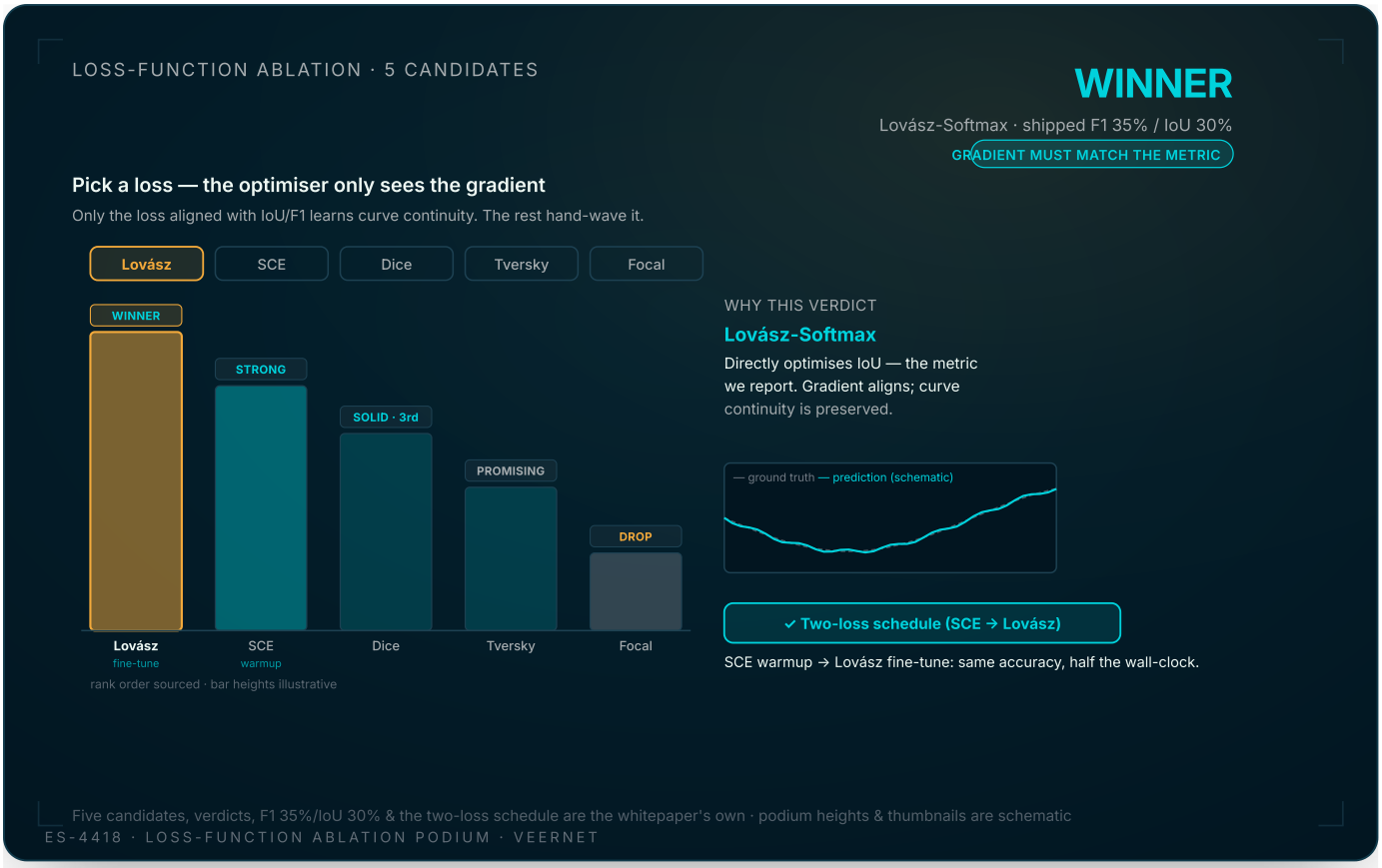
Ablations: the evidence that the design choices are load-bearing

A whitepaper that reports only its best configuration is a brochure. The case for GeoBFDT rests on ablations that show each unusual choice earning its place — and several of them are counter to the instinct of a team coming from natural-image detection.

Backbone: smaller is dramatically better. The instinct is to reach for ResNet-18 or ResNet-50. On this small-data regime that instinct is wrong by an order of magnitude. Swept across ResNet-10, -14, -18, and -34 with everything else held fixed, the **ResNet-10 backbone reaches a classification error of about 0.5%**, while ResNet-18 lands near 21% and ResNet-34 near 27%. The deeper backbones overfit 14 wells; the lightweight one generalises. This is the single most important — and most surprising — engineering finding in the programme.

Augmentation, applied where it counts. Without augmentation the model collapses to constant prediction: classification error of 100%. With a sinusoid-aware augmentation recipe applied to the feature-bearing patches — colour jitter, sharpness, blur, and noise variants — **classification error falls from 100% to about 2.6%**. Augmentation here is not a generic regulariser; it is the difference between a model that learns nothing and a model that works.

Dynamic versus static imagery, and well count. Training on the dynamically-normalised image rather than the static one moves classification error from about 63% to 2.5%. And the data itself is non-stationary: stepping the training set from 3 wells to 14 collapses classification error from roughly 93% toward the low single digits, the clearest possible argument that the bottleneck on this problem is labelled wells, not model capacity.



Loss-function choice decides whether the network learns curve continuity. VeerNet tested five losses under identical conditions; only the one whose gradient aligns with the IoU/F1 metric (Lovász-Softmax) wins, and the shipped answer is a two-loss SCE-warmup → Lovász-finetune schedule. Pick a loss to see its ablation verdict, the reason, and a schematic ground-truth-vs-prediction trace; toggle the two-loss schedule (same accuracy, half the wall-clock). The five candidates, verdicts, F1 35%/IoU 30% and the two-loss schedule are the whitepaper's own; the podium bar heights are ordinal (rank sourced) and the prediction thumbnails are schematic.

The podium above ranks these interventions by the loss they remove. The ordering is the message: the choices that matter most are not exotic architectural tricks but disciplined data and backbone decisions — sinusoid-aware augmentation, a deliberately small backbone, dynamic input, and more labelled wells. Each is a tracked, reproducible experiment, not a remembered preference, and each is the kind of decision that has to be defensible a year later when an operator's own team retrains on new wells.

The engineering that makes it production-grade

The model is the visible artefact; the pipeline around it is what makes the numbers above reproducible and the system ownable. Three engineering disciplines deserve naming, because they are what separate a paper result from a deployable one.

Data engineering against a brutal raw format. Each well image is enormous — on the order of 690,000 by 360 pixels and roughly 1.5 GB — so the pipeline cuts overlapping patches at a fixed stride (40 pixels) and a fixed patch height of **800 pixels, chosen because it corresponds to 2.2 m and contains over 95% of fracture and bedding sinusoids in full**. Image-log data also arrives dirty: in one intake of ten wells, two were excluded before training because their static-image value ranges fell wildly outside the normal band and defeated normalisation, with one of the two additionally acquired on a different imaging tool whose response was not directly comparable. That exclusion is a versioned data decision with model consequences, not a silent edit — the kind of provenance that lets a reviewer reconstruct exactly which wells produced a given number.

A reproducible training configuration. The converged recipe is a single readable record: ResNet-10 from scratch, 4+4 encoder/decoder, feed-forward dimension 1024, AdamW at an optimal learning rate of 0.0004, dropout 0.2, batch size 128, focal-plus-L1 loss with class/parameter weights of 5 and 1, inference threshold 0.5, and early stopping after 40 epochs without improvement. That is not folklore; it is a pinned artefact an operator's engineers can read off and retrain from.

A CV pipeline, not a notebook. GeoBFDT is one stage in a computer-vision system — ingestion of the binary image log, normalisation, patch generation, augmentation, set-prediction inference, and conversion of normalised query outputs back into physical depth, dip, and azimuth — built to run inside the operator's perimeter rather than as a researcher's local script. The discipline that matters is that every transformation between the raw binary log and the final pick is explicit and versioned, because a dip prediction feeds a structural model that feeds a drilling decision, and "the network said so" is not an answer a reviewer can accept.

Where this belongs in a geoscience workflow

GeoBFDT is not a replacement for the structural geologist; it is a force multiplier that changes what the geologist spends time on. The model does the mechanical first pass — emitting candidate fractures and beddings with their geometry across kilometres of image log — and the expert is retained for the anomalies, the ambiguous intervals, and the sign-off. The geometry accuracy is high enough (dip near 90% at 3 degrees, azimuth near 92% at 15 degrees) that the picks are usable inputs to a structural model rather than a rough

screen, and the set-prediction design means the crowded, intersecting intervals — the ones that defeat both manual picking throughput and mask-based detectors — are exactly where the model adds the most.

For an R&D geoscience lead evaluating this class of model, the decision criteria are concrete:

- Does the detector output the geological parameters you actually consume (depth, dip, azimuth) directly, or a mask you then have to fit?
- Can it represent multiple, overlapping features in a single patch, or is it structurally one-feature-per-image?
- Is every reported metric tied to an explicit localisation tolerance read against the instrument's physical resolution floor?
- Is the training configuration a pinned, reproducible artefact your own team can retrain from on new wells?
- Are the design choices backed by ablations, or asserted?

GeoBFDT was built — across our engagements with operators in the Middle East and the United States, and validated in depth on the Middle East carbonate dataset we partnered on — to answer yes to all five. The architecture is, in the end, a claim about inductive bias: that the right way to find fractures in a fractured carbonate is not to paint better masks, but to ask the model for a set and to score the set as a whole. The ablations are the evidence that the claim holds.

WHAT THIS WHITEPAPER ARGUES

1. Fracture and bedding picking in fractured carbonates is a set-prediction problem, not a per-feature segmentation problem — multiple sinusoids overlap and intersect in a single image patch, and mask-based, one-feature-per-image detectors plus NMS structurally cannot handle that.
2. GeoBFDT is an end-to-end Detection Transformer: a ResNet-10 backbone trained from scratch, a 4+4 transformer encoder/decoder, learned object queries, and a Hungarian bipartite-matching loss (focal + L1) that regresses depth, dip and azimuth directly — no masks, no anchors, no post-processing.
3. On 14 vertical wells (two different microresistivity imaging tools) from a Middle East carbonate play: F1 ~65%/63% (fractures/beddings) at 3 cm tolerance and ~75%/69% at 5 cm, with dip accuracy ~90% at 3 degrees and azimuth ~92% at 15 degrees — every detection metric read against the ~3 cm physical floor set by the digital log format's pixel resolution.
4. Ablations are load-bearing: a ResNet-10 backbone beats ResNet-18/34 by an order of magnitude on classification error in this small-data regime, sinusoid-aware augmentation moves classification error from 100% to ~2.6%, dynamic input beats static, and more labelled wells (3 to 14) is the dominant lever.
5. The production case rests on the engineering around the model — versioned data with recorded QC exclusions, a pinned reproducible training recipe, and an in-perimeter CV pipeline — not on the geoscience alone.

References

Carion et al., 2020 N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko. End-to-End Object Detection with Transformers (DETR). ECCV 2020. The set-prediction and bipartite-matching formulation GeoBFDT is built on. https://arxiv.org/abs/2005.12872

He et al., 2017 K. He, G. Gkioxari, P. Dollar, R. Girshick. Mask R-CNN. ICCV 2017. The mask-based instance-detection paradigm the borehole-image baselines descend from. https://arxiv.org/abs/1703.06870

Kuhn, 1955 H. W. Kuhn. The Hungarian Method for the Assignment Problem. Naval Research Logistics Quarterly, 1955. The polynomial-time bipartite-matching algorithm at the core of the GeoBFDT loss. https://onlinelibrary.wiley.com/doi/10.1002/nav.3800020109

Lin et al., 2017 T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar. Focal Loss for Dense Object Detection. ICCV 2017. The classification loss that handles the dominant no-object class in set prediction. https://arxiv.org/abs/1708.02002

He et al., 2016 K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. CVPR 2016. The residual backbone family from which the ResNet-10 image encoder is drawn. https://arxiv.org/abs/1512.03385

GeoBFDT: End-to-End Detection Transformers for Fracture and Bedding Picking in Carbonate Image Logs

Published May 2024. Generated 2026-07-22.
https://www.earthscan.io/whitepapers/geobfdt-end-to-end-transformers-fracture-bedding-carbonate