

Online Resource 2

Execution provenance and reproducibility record.

Bounded Repair Reachability of Residual Critical Blocks in 3D Binary Masks under
SAFE-Gated Single-Voxel Flips

Journal of Mathematical Imaging and Vision

Authors

Jihun Bae; Yeonho Bae; Jinglu Hu

Corresponding author

Jihun Bae
Graduate School of Information, Production, and Systems, Waseda University
Kitakyushu, Japan
jhbae1184@akane.waseda.jp

Online Resource 2 (file title: Supplementary Note 2). Execution provenance and reproducibility record

Jihun Bae, Yeonho Bae and Jinglu Hu

1 Scope

This note records how the searches of the main text were produced, reused and checked. It makes no claim beyond the main text.

2 Lineage of the starting states

Table OR2.1. Starting-state lineage and code provenance.

Item	Origin	Identifier
Production implementation of R^+ (<code>r_plus_repair_production.py</code>)	authors' earlier unpublished work, code base; imported unchanged by the SAFE-gated Stage 1	SHA-256 878b483f300415f6a3bde2169ce6630de890842c8a79326d99a0f50a2ecb255b
<code>cubical_betti.py</code>	authors' earlier unpublished work, code base	SHA-256 d87571d09f828288a9660c65b6b841d4f7939f70d9c9b1b082bfc99bde5a5b64
RAW masks, recorded R^+ edits (876 cases)	authors' earlier unpublished work, materialization archive	archive SHA-256 prefix 784fc692...; its freeze record carries the status <code>LOCAL_EVIDENCE_FROZEN</code> and <code>canonical_lineage_ready = false</code> , which records that the administrative freeze of the inherited P3 data-environment archive is not attached; the record states that this gate does not alter the case selection or the arrays
Gate generator <code>exhaustive_gate.cpp</code>	P0 audit of this study	SHA-256 fa1a3fd564c553c8fe2a02eb9803197cb428a42bca96c64a10aeef0bb7a8a7c0
Gate table	regenerated from the generator	SHA-256 59407fef2acb5b06059a13f19092047d172bbddd49bc85f811d822293b4c4830

Stage 1 differs from R^+ in one selection step (SAFE filter); Stage 2 adds SAFE removals. Stage 1 imports the queue, the block detection and the update of incident blocks of the production R^+ unchanged; this note does not re-test that equivalence case by case.

3 Look-ahead engine

The earlier searches used a look-ahead engine that stores the breadth-first levels 0 to $L - 2$ exactly as the production search does. For each state of level $L - 2$ it generates the children (level $L - 1$) without storing them. A child is discarded only if a necessary condition for reaching a success state within one further flip fails. When the flip lay outside B and B was critical in the parent, B is still critical in the child, and some voxel of B must then admit a SAFE flip in the child state (the gate is evaluated on the 26-neighbourhoods of the child, not on those of X_T or of the parent, and under F2 the flip must not remove input foreground); for F1, $\text{NRFR}(\text{child}) \leq 1$. A child that is a success state ends the search at depth $L - 1$. Otherwise, the last flip is tested only for the voxels that lie in every critical block counted against the success criterion (B and any critical block absent from X_T) and, for F1 when exactly one input-foreground voxel is missing, for the restoration of that voxel; when no such block is critical in the child, all admissible voxels qualify. Each candidate is tested directly. A last flip outside this candidate set cannot produce a success state, because a flip changes the status only of the blocks containing the flipped voxel and changes NRFR by at most one (Online Resource 4, completeness argument). The engine reports "no success state within L flips" and does not separate EXH from DL, so its negative verdicts are counted as DL; the statement they carry, absence of a success path of at most L flips for the system, window and gate, is the same in both cases.

4 Standing of the inherited verdicts

These inherited absence verdicts are of a different standing from the verdicts computed in the production run, and they are labelled as such. Their completeness rests on the argument above, on the audit of the engine code against it, and on comparisons of the engine with the unchanged breadth-first search on the rows on which the latter finishes (162 rows of the earlier, unchanged search with 0 differences, among them 27 F1 repairs, 12 of which end with the re-addition of a voxel removed earlier on the path, as in the repair of BraTS2021_00110; 125 rows of the engine run with 0 differences). Of the 125 rows of the engine run compared with the unchanged search, 80 are among the 611 inherited DL rows (45 in C00-F1, 35 in C00-F2; the unchanged search was run at $R = 2$ and $L = 6$ with a raised state limit) and 5 are among the 44 inherited FOUND rows; all agree. These rows were selected by projected state count and are the smaller ones, so the agreement does not extend by itself to the other 531 DL rows, which were not reproduced in that comparison. No second implementation of the engine exists. The FOUND rows were verified in addition by replaying all 45 witnesses of the engine run on the full volume with an independent step-by-step implementation of the gate and the system rules. Statements in Sections 4 and 5 that depend on inherited absence verdicts say so and give the number of rows that rely on them.

5 Pilot verification (G1)

Pilot verification of 24 inherited rows (G1). Of 24 further inherited DL rows examined in a later pilot, 12 were re-decided at the original horizon $L = 6$ and 12 were re-evaluated at the reduced horizon $L = 5$ only. All 24 were searched with the unchanged breadth-first search ($R = 2$, state limit 2×10^7 , one process at a time). The selection rule, the search modules and the gate table were fixed by SHA-256 digest before the first row was run. The 24 rows (24 blocks in 19 cases) were drawn from the 531 inherited DL rows that were not among the 80 re-decided above, in the cells $\text{C00-F1} \times \text{C1}$, $\text{C00-F1} \times \text{C2}$ and $\text{C00-F2} \times \text{C1}$, at the minimum and at the 25th, 50th and 75th percentiles of the projected cumulative number of states. Rows were eligible only if the projected number of states to the compared depth did not exceed 1.2×10^7 , which leaves 392 of the 531 rows eligible and excludes the other 139. Twelve rows (full-depth rows) were searched to the original horizon $L = 6$ and twelve (reduced-depth rows) to $L = 5$, in each case beyond the last level (4 or 5) that the earlier frozen search of the row had completed before reaching its state limit. For the reduced-depth rows the look-ahead engine was also run at $L = 5$. All 24 searches completed below the state limit, and none found a success state. The full-depth searches enumerated every state within six

flips (2.03×10^6 to 1.18×10^7 states per row) and the reduced-depth searches every state within five flips (1.49×10^6 to 8.15×10^6). For all 24 rows the numbers of states at levels 0 to 4 equal those stored by the engine, and for the 12 reduced-depth rows the engine at $L = 5$ and the search agree that no success state exists. The 12 full-depth rows re-decide the inherited verdict that no repair of at most six flips exists. The 12 reduced-depth rows establish the absence of a repair of at most five flips and the agreement of the engine with the search at that horizon; they do not re-decide the six-flip verdict. The counts 80 and 531 above are the record of the earlier comparison and are not revised by the pilot. Counting the 12 full-depth rows, 92 of the 611 inherited DL rows have been re-decided at $L = 6$ and 519 have not. The sample is drawn from the rows with smaller projected state counts and supports no estimate of the agreement rate over all 531 rows.

6 Provenance

Provenance. The production run and its checks use the frozen search modules, imported unchanged and identified by SHA-256 digest, and the gate table pinned by digest `59407fef...4830`. The digest is that of the regenerated table; the original record of the table is not available. The state X_T of each case is reconstructed from the pinned RAW masks and the recorded edit sequence of stage 2, and its mask digest must equal the recorded one (222 of 222 cases). The production run used Python 3.13.16, numpy 2.5.3, scipy 1.18.1 and nibabel 5.4.2 (versions recorded in the shard metadata and in the pre-run freeze record), PYTHONHASHSEED=0 and two shards. The look-ahead engine was run with Python 3.13.16 and numpy 2.5.3; the earlier searches that were not repeated (Section 4.1, Table 4) ran with Python 3.13.15. Searches are deterministic: a rerun of a shard gives the same rows apart from timing columns, and the result does not depend on how the targets are divided into shards. A dedicated verifier replayed every FOUND witness of the production run on the full volume, step by step, with an explicit per-voxel gate and the rules of the system, and re-ran sampled rows with an independent breadth-first search.

7 Re-aggregation

Re-aggregation. The counts in the tables and the numbers quoted in the text were recomputed from the production tables `BLOCKS.csv` and `PATHS.csv`, and the reuse log by one read-only script (`reaggregate_r1_2.py`); three stored witness bundles of BraTS2021_00110 are optional inputs. Before any aggregation the script compares the SHA-256 digest of every input with a master list. It stops with a nonzero exit code if a required input is missing, if the bundles are present only in part, if a file is not listed or differs from the list, if a required column or array is missing, or if a `PATHS` record fails validation (move type, 0/1 flags, SAFE flag equal to 1, and steps $0 \dots L^* - 1$ for every FOUND row). Ten defect cases of this kind were tested with the preceding version `reaggregate_r1_1.py`, and each ended with the intended exit code; `reaggregate_r1_2.py` differs from it only in making the bundle checks optional (patch script, tests and logs in Online Resource 4). Without the bundles the checks of the witness of BraTS2021_00110 that read the stored arrays (NRFR along the witness, stored `lCritl` after each step, SAFE flags of the window of X_T) are reported as skipped; all other checks and all tables are identical with and without the bundles. *Recomputed* values are derived by the script from stored rows or arrays: the status counts, the DL and CAP invariants (in all 2,760 DL rows fewer than $N_{\max}$ states were enumerated; all 919 CAP rows stopped at exactly $N_{\max}$), the cohort splits, the comparisons of paths between cells and the NRFR along the witness of BraTS2021_00110. *Inherited* values were written by the production run or by earlier audits and are only read, counted or compared: the SAFE flags of recorded moves and window voxels, the numbers of critical blocks after each step of the witness, the per-witness counters `net_raw_fg_removed`, `rawfg_rem_steps` and `n_repeated_flip_voxels`, the prior labels of the 655 reused rows, the start-state qualification of the 680 blocks, and the figure audit. No search is re-run, and neither the search engine nor the SAFE gate is re-implemented. The script shows that the stored results agree with the tables reported here; it is not an independent verification of the searches.

8 Split check

Split check. A second script (`check_split_r1_1.py`) compares the 876 input identifiers with the TRAIN list of the split contract. The two lists are equal as sets, the 222 cases with target blocks lie in the list, and the stored digests of all 1,251 eligible cases equal the SHA-256 digest of the salt followed by the identifier. The split contract archive has SHA-256 `1202b259801b3f3448c3e04448d5222ce2cb7e77ee31d16e564b33fabb52ba11`.

9 Reproduction of the SAFE table and comparison with the topological-number criterion

The frozen generator was compiled unchanged (g++ 11.4.0, -O2) in a scratch directory and run twice. Its source digest equals the recorded one. Both runs wrote tables of 67,108,864 bytes with SHA-256 `59407fef...4830` (identical to the pinned table) and exit code 0. Reported: 67,108,864 patterns; 15,935 distinct attachments; configurations accepted by the Betti condition, by the topological-number criterion ($T_{26} = 1$, $\bar{T}_6 = 1$; its correspondence with the definition in [BM94] has not been checked against the original) and by collapsibility: 25,985,144 each; disagreements 0; collapse-order-dependent attachments 0. Every table entry has all three decision bits equal (counts 41,123,720 with none set and 25,985,144 with all set). Among the Betti classes, only $(h_0, h_1, h_2) = (1, 0, 0)$ contains accepted configurations. Limits of this check: the three decisions are computed inside one program, with the topological-number branch written separately from the attachment branch; the check reproduces a recorded result and is not an independent implementation; it concerns neighbourhood configurations and not whole images. The independent boundary-matrix computation of Online Resource 1 is a separate check of the Betti branch. An independent recount (`simple26.c`) was written separately from the generator; it can be run without a table, in which case it does not read one. It counts the 26-connected components of the foreground in N_{26} and the 6-connected components of the background in N_{18} that contain a face neighbour (topological numbers) and, in a second branch, the Euler characteristic and connectedness of the union of the faces of C_v shared with the foreground neighbour cubes, which is equivalent to the Betti condition by Online Resource 1, Proposition 1. Over all 67,108,864 configurations both branches accept 25,985,144 configurations (38.7 %), they never differ, and the attachments take 15,935 distinct values. Run against the table written by the regenerated generator (SHA-256 `59407fef...4830`), the attachment branch equals bit 0 and bit 2 (collapsibility) and the topological-number branch equals bit 1 of every one of the 67,108,864 entries (0 entry disagreements; all three bits set in 25,985,144 entries and none in the other 41,123,720). Copies of the table in which one entry has bit 0, bit 1 or bit 2 altered are rejected with exit code 3, and a table one byte short is rejected with exit code 1. Compiled with `gcc -O2 -Wall -Wextra -Werror` without diagnostics; one run takes five to eight seconds on the machines used.

10 Blocks that are FOUND at G14 and CAP at G26 (C00-F0)

Ten blocks: BraTS2021_00104 k2xy@88,105,115, BraTS2021_00457 k2xy@57,144,144, BraTS2021_01228 k2zx@49,122,104, BraTS2021_01248 k2zx@81,143,172, BraTS2021_01267 k2xy@66,177,136, BraTS2021_01267 k2yz@66,176,136, BraTS2021_01267 k2yz@66,177,136, BraTS2021_01267 k2zx@66,177,136, BraTS2021_01293 k2xy@72,71,152, BraTS2021_01293 k2yz@72,71,152. In the effective counts, no block of C00-F1 or C00-F2 shows this pattern. The raw production table is used for C00-F0, which has no substituted rows.

11 Structural and observed zeros at G26

EXH after the initial state only (no admissible first move): C10 and C11P, 595 of 680 blocks in each criterion; C11H, 680; C00 under F2, 69; C01P under F2, 69; C01H under F2, 103. All other zero FOUND entries come from searches that enumerated more than one state (C10 and C11P: 85 blocks each).

12 Online Resource 4 inventory

Data availability. The BraTS 2021 challenge data (NIfTI images and segmentations) are listed by The Cancer Imaging Archive under CC BY 4.0 with a required data citation [BraTS21data] (page read on 2026-10-11) and are available from the challenge organisers [Baid21; Menze15; Bakas17]; they are not redistributed. The original DICOM images of the source collections are listed there under the NIH Controlled Data Access Policy and a TCIA usage agreement, and this study does not use them. The split is reproducible from the list of eligible case identifiers and the salt, which are part of the Online Resource 4. The Online Resource 4 (derived results) comprise Tables S1–S3, the production tables `BLOCKS.csv` and `PATHS.csv`, the reuse log, the distribution hash list (five lines, a subset of the preserved master list), the re-aggregation script `reaggregate_r1_2.py` with its patch script and tests, the split-check and split-regeneration scripts with their outputs and negative tests, the comparison of the block rule with the definition of Section 3.2, the completeness argument, tests and comparison logs of the look-ahead engine, the selection table, results table, scripts, pre-run hash record and journal of the G1 pilot verification, and the TRAIN identifier list, the production runner (`p2ab_ablate.py`) and the frozen search modules (`p2ab_search.py`, `P2B_search_R1.py`, `p2ab_complete.py`), the generator of the SAFE table (`exhaustive_gate.cpp`), the independent recount with its table-comparison and tamper-test scripts (`simple26.c`, `run_safe_check.sh`, `tamper_table.py`), Online Resource 1 with the programs and logs that verify its finite computations (`s2_lemma_check.py`, `safe_independent_check.py`), and the versions and SHA-256 hashes of the code, the table and the software environment (Python 3.13.16, numpy 2.5.3, scipy 1.18.1, nibabel 5.4.2). They contain case identifiers, block coordinates and search results, and no image intensities. Specifications and reports originally written in Korean are given in English translation (files `*_EN.md`); `REDACTIONS.csv` lists every file that was withheld, translated or edited, `DEPENDENCIES_ICEIC27.md` the unpublished modules that the searches require, and `README.md` the reproducibility levels. The cropped mask windows used for main-text Figure 7 and Online Resource 3, Figure S1 are not distributed. The window coordinates are included. Online Resource 4 retains the recorded-result and verification materials described above. The versioned public software and its selected-search scope are identified in the main manuscript's Code availability statement. With a lawfully obtained BraTS2021_00110 input matching the canonical-array hash in `DATA_ACCESS.md`, that release reconstructs the selected starting state by replaying frozen ordered edits and executes four core searches. It does not rerun the cohort-wide starting-state generation stages or all 19,040 searches. The historical production environment recorded above is distinct from the selected-case release environment documented in its `README` and validation record.