

Reading status. The detailed review below preserves the earlier literature synthesis and its proposed ARIA-NBV adaptations. Recommendations about model inputs, training stages or implementation are research proposals, not a description of the current system or evidence of measured benefit. The active thesis and implementation specify the current choices.

SCONE And FisherRF

Primary sources. SCONE [1] and FisherRF [2].

External pointers. SCONE: arXiv 2208.10449, NeurIPS 2022 abstract. FisherRF: arXiv 2311.17874. See also NeRF [3], 3D Gaussian Splatting [4], and the e3nn spherical-harmonics reference [5] for the representation families touched by these papers.

Local source mirrors. SCONE source: `camera_ready.tex`, `camera_ready_2_approach.tex`, and `camera_ready_6_appendix.tex`. FisherRF source: `main.tex`, `sec/method.tex`, and `sec/exps.tex`. The manifest entries live in `sources.jsonl`.

Related ARIA-NBV pages. RRI theory, active 3DGS and targeted NBV, RL planning, and target-aware thesis questions.

Executive Takeaway

SCONE and FisherRF are useful for ARIA-NBV, but not as replacement objectives.

- SCONE contributes a principled coverage-support view: infer where unseen surface likely exists, estimate whether a candidate can see it, and integrate visibility over proxy volume points.
- FisherRF contributes a principled information view: score views by expected information gain about a scene representation, using Fisher information rather than only heuristic uncertainty maps.
- ARIA-NBV should use both as actor-visible auxiliary channels for candidate tokens, candidate diagnostics, rollout branch diversity, and support-coverage checks.
- The thesis utility remains target-specific Relative Reconstruction Improvement (RRI) and endpoint target reconstruction quality. Coverage, uncertainty, semantics, simulators, and continuous-control policies stay bridges or diagnostics until calibrated against target-RRI.

The shortest thesis-safe synthesis is:

Target-RRI is the label and evaluation target,
SCONE/FisherRF provide support and information features.

Conceptual Comparison

dimension	SCONE	FisherRF	ARIA-NBV usage
Primary quantity	Surface coverage gain.	Model-parameter information gain.	Target-RRI remains primary.
Scene representation	Probabilistic occupancy plus proxy points.	Radiance-field parameters and Hessian/Fisher information.	EVL, semi-dense points, target crops, and support cells.
View value	Expected newly visible surface.	Expected entropy reduction.	Auxiliary support/information features and diagnostics.
Unknown geometry handling	Occupancy probability plus Monte Carlo integration.	Parameter uncertainty plus rendering Jacobians.	Actor-visible support proxies, not oracle future labels.
History representation	Camera history around proxy points on a sphere.	Accumulated information matrix.	$\mathbb{S}^2$ directional memory plus support counts.
Main risk	Coverage can miss reconstruction quality.	Information gain can miss geometry quality.	Never replace target-RRI.

The unifying distinction is:

SCONE : $\mathbb{E}[\text{new visible surface}]$,
 FisherRF : $\mathbb{E}[\text{uncertainty reduction}]$,
 ARIA-NBV : $\mathbb{E}[\text{target reconstruction-quality improvement}]$.

SCONE: Surface Coverage As Support

SCONE starts from the classic NBV coverage goal: choose a pose q that observes the largest amount of previously unseen surface. If the true surface S were known, the gain could be written as:

$$G_{\text{surf}}(q) = \int_{x \in S} \mathbf{1}[x \text{ visible from } q] \mathbf{1}[x \text{ not previously observed}] dA(x).$$

The problem is that S is unknown. Extracting a surface from a probabilistic occupancy field is brittle because a surface is a zero-measure set in a 3D volume. SCONE’s theoretical move is to replace the surface integral with a volumetric integral over a thin neighborhood around the surface. Its source derives that, under regularity assumptions and for small neighborhood thickness μ , the volume integral is asymptotically proportional to the surface coverage gain:

$$\left| \frac{1}{|\chi|_V} \int_{\chi} g_c^H(\mu; x) dx - \mu \frac{|\partial \chi|_S}{|\chi|_V} G_H(c) \right| \leq M\mu^2.$$

Conceptually:

$$\text{surface coverage} \approx \text{volume integral of occupancy-weighted visibility gain}.$$

SCONE therefore asks which sampled volume points are likely near an unobserved surface and visible from a candidate pose, rather than requiring a clean mesh first.

SCONE Architecture Details

SCONE has two learned modules that should be understood as inspiration, not thesis-core dependencies.

module	paper mechanism	ARIA-NBV transfer
Occupancy probability	A deep implicit function predicts occupancy from the partial point cloud and query point features. It uses local multi-scale neighborhoods, attention over centered neighbor sequences, pooling, and spatial cells for scalability.	Replace full-scene learned occupancy with target-local support probability from EVL, semi-dense points, predicted/observed target crops, and support counts.
Visibility gain	Proxy points are sampled from predicted occupancy, filtered by candidate frustum, encoded with occupancy and coordinates, processed with self-attention for occlusion effects, augmented with camera-history features on a sphere, and decoded into spherical-harmonic visibility coefficients.	Use candidate-frustum, ray/depth, and directional-memory features as transparent candidate support channels.

The ARIA-relevant detail is the camera-history encoding around each proxy point. SCONE represents where previous cameras were relative to a point on a sphere. ARIA-NBV can use a simpler second-moment directional memory:

$$d_k(v) = \frac{c_k - v}{\|c_k - v\|}, \quad M_{\text{dir}}(v) = \sum_{k < t} w_k(v) d_k(v) d_k(v)^\top.$$

This supports the design rule that view history should be represented directionally around target-local points, not only as a sequence of global poses.

Figure 1 separates the occupancy estimate from the directional visibility prediction.

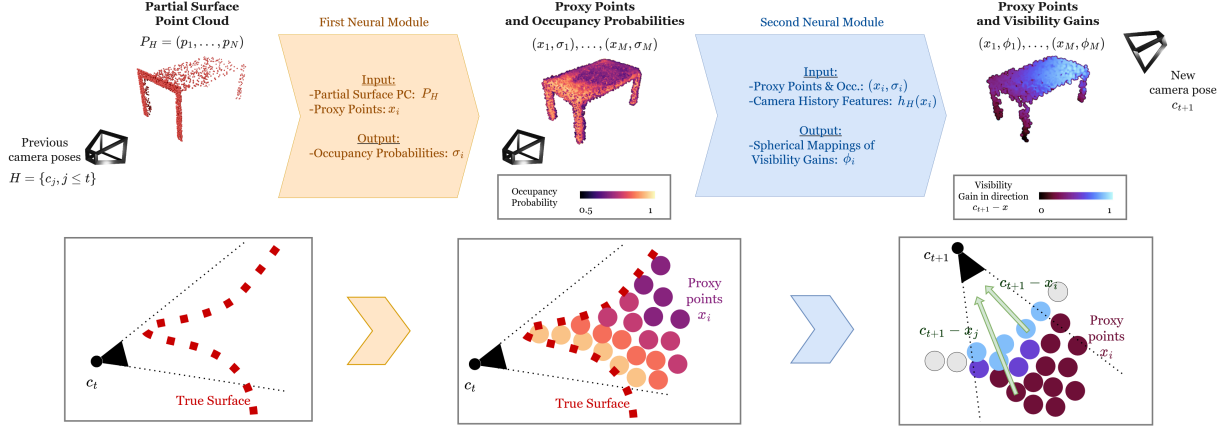


Figure 1: Scone first estimates occupancy at proxy points, then predicts their directional visibility gain. Reproduced from [1, Fig. 2].

SCONE Transfer To ARIA-NBV

The target-local analogue of Scone's surface integral is:

$$G_{\text{target-vis}}(q, e) = \int_{v \in \Omega_e} p_{\text{target-surface}}(v) p_{\text{visible}}(v, q) p_{\text{not-yet-observed}}(v) dv.$$

Here Ω_e is the actor-visible target region, such as an observed or predicted Oriented Bounding Box (OBB) crop. The target-surface term can come from EVL, semi-dense support, object confidence, and target support counts; visibility can come from frustum/ray/depth approximations; and not-yet-observed evidence can come from $\mathbb{S}^2$ directional memory.

A direct Monte Carlo candidate feature is:

$$u_{\text{SCONE}}(q, e) = \frac{1}{M} \sum_{m=1}^M \hat{p}_{\text{surf}}(v_m) \hat{p}_{\text{vis}}(v_m, q) \hat{n}_{\text{dir}}(v_m, q), \quad v_m \sim \Omega_e.$$

A more factual, implementation-first support integral is:

$$\hat{U}(q, e) = \frac{1}{M} \sum_{v_m \in \Omega_e} w(v_m) \mathbf{1}[v_m \in \text{Frustum}(q)] \mathbf{1}[\text{ray not blocked}] \mathbf{1}[\text{direction novel}].$$

This is valuable because the failure modes are inspectable:

- candidate cannot see target support,
- target crop is empty,
- candidate support is too low,
- direction is not novel,
- free-space or collision validity failed,
- learned scorer disagrees with oracle target-RRI.

SCONE also suggests a target support probability rather than full-scene occupancy:

$$\hat{p}_{\text{target-surface}}(v) = f_{\text{support}}(F_0^{\text{EVL}}(v), P_t(v), \hat{B}_e, \hat{y}_e).$$

For the thesis core, this can start as a transparent hand-built proxy:

$$\hat{p}_{\text{target-surface}}(v) = \mathbf{1}[v \in \hat{B}_e] \text{clip}(\alpha \text{EVL-support}(v) + \beta \text{semidense-support}(v), 0, 1).$$

A learned target-local occupancy/support head is a later ablation, not an M1-M3 prerequisite.

Support Coverage Before Q_H

SCONE reinforces the support-coverage gate before interpreting a learned value model. Before comparing one-step scorers, oracle lookahead, or Q_H , ARIA-NBV should know whether the finite candidate set contains enough valid target/candidate support.

Report candidate support by family:

$$\text{valid support}(q, e) = \sum_{v \in \Omega_e} \hat{p}_{\text{surf}}(v) \hat{p}_{\text{vis}}(v, q).$$

The relevant strata are:

- forward_local
- target_bearing_local
- lateral_target_bypass
- upper_bound_free_shell

If upper_bound_free_shell has high target support but realistic families do not, the bottleneck is candidate generation, not Q_H .

FisherRF: Information Gain For Views

FisherRF motivates view selection through expected information gain about a radiance-field model’s parameters and selects candidates using a tractable trace surrogate. It is relevant because it provides a principled uncertainty-reduction channel, not because ARIA-NBV should train a radiance field as its core scene state.

Figure 2 shows the mapping loop used in the paper: candidate paths are evaluated against the current reconstruction.

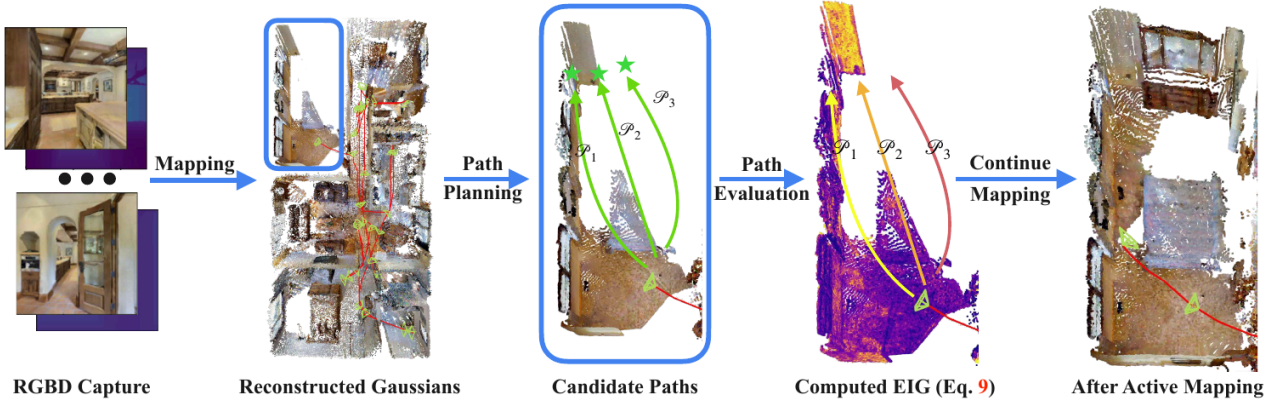


Figure 2: FisherRF mapping example: acquired RGB-D observations update Gaussians, and predicted information gain compares candidate paths. The depicted reconstruction belongs to the source paper. Reproduced from [2, Fig. 2], with surrounding whitespace removed.

For model parameters θ and candidate observation y_q , Fisher information is:

$$\mathcal{I}_q(\theta) = \mathbb{E}_{y_q} [\nabla_{\theta} \log p(y_q | q, \theta) \nabla_{\theta} \log p(y_q | q, \theta)^{\top}].$$

Under regularity conditions, Fisher information equals the expected negative log-likelihood Hessian. FisherRF uses a rendering-Jacobian expression with assumed unit RGB measurement covariance:

$$H''[y | x, \theta] = \nabla_{\theta} f(x; \theta)^{\top} \nabla_{\theta} f(x; \theta).$$

The key point for active view selection is that Fisher information does not require the unknown future RGB observation. It depends on how sensitive the rendered observation would be to scene parameters.

If F_t is the information accumulated from existing views and F_q is the candidate information, expected information gain can be written as entropy reduction:

$$\text{EIG}(q) = H(\theta | D_t) - H(\theta | D_t \cup q).$$

Under a Gaussian/Laplace approximation:

$$\text{EIG}(q) \approx \frac{1}{2} \log \det(F_t + F_q + \Lambda) - \frac{1}{2} \log \det(F_t + \Lambda).$$

Because full radiance-field Hessians are enormous, FisherRF uses a diagonal/Laplace-style approximation plus regularization. The diagonal intuition is:

$$\text{EIG}(q) \approx \frac{1}{2} \sum_j \log \left(1 + \frac{f_{q,j}}{f_{t,j} + \lambda} \right).$$

These log-determinant expressions describe the information-gain approximation. FisherRF’s view-selection rule maximizes its trace upper bound, using a regularized diagonal approximation to the rendering curvature [2].

This information-gain intuition says a candidate is valuable when it provides information about parameters or regions that are still uncertain, not when it repeats already well-constrained evidence.

FisherRF Transfer To ARIA-NBV

The following target-local scores are proposed ARIA-NBV transfer ideas inspired by the information-gain motivation. Their value for target-RRI ranking remains to be measured.

ARIA-NBV should not ask “what RGB image will this candidate observe?” as an oracle input. The Fisher-style actor-visible question is:

Which uncertain target-local variables would this candidate constrain?

Candidate-constrained variables can include:

- target occupancy,
- target surface support,
- target OBB geometry,
- target completeness,
- target-local directional visibility,
- EVL occupancy/evidence logits,
- semi-dense support gaps,
- learned target-quality latent features.

A target-local Fisher-like channel is:

$$u_{\text{Fisher}}(q, e) = \sum_{v \in \Omega_e} \log \left(1 + \frac{f_q(v)}{f_t(v) + \lambda} \right),$$

where $f_t(v)$ is current information/support and $f_q(v)$ is candidate-provided information/support. A simple non-differentiable ARIA version is:

$$f_t(v) = n_t(v) + \lambda_{\text{dir}} \text{tr } M_{\text{dir}}(v), \quad f_q(v) = \mathbf{1}[v \in \text{Frustum}(q)] \hat{p}_{\text{visible}}(v, q) \hat{p}_{\text{target - surface}}(v).$$

Then:

$$u_{\text{info}}(q, e) = \sum_{v \in \Omega_e} \log \left(1 + \frac{\hat{p}_{\text{visible}}(v, q) \hat{p}_{\text{surf}}(v)}{n_t(v) + \lambda_{\text{dir}} \text{tr } M_{\text{dir}}(v) + \lambda} \right).$$

This is a target-local uncertainty-reduction feature. It should not replace RRI, but it can become a candidate-token feature for the one-step scorer and Q_H .

Directional Novelty And Batch Diversity

SCONE’s spherical history and FisherRF’s diminishing returns combine into a concrete directional novelty term:

$$\hat{p}_{\text{unseen - dir}}(v, q_i) = 1 - \frac{d_i(v)^\top M_{\text{dir}}(v) d_i(v)}{\text{tr } M_{\text{dir}}(v) + \epsilon}.$$

Use it in the support channel:

$$u_{\text{vis}}(q_i, e) = \sum_{v \in \Omega_e} \hat{p}_{\text{surf}}(v) \hat{p}_{\text{frustum}}(v, q_i) \hat{p}_{\text{unseen - dir}}(v, q_i).$$

FisherRF also gives a cleaner branch-diversity rule than only sibling distance/yaw guards. Greedily select rollout branches by marginal information:

$$q_1 = \arg \max_q u(q), \quad q_k = \arg \max_q \Delta u(q \mid q_1, \dots, q_{k-1}).$$

In diagonal target-support form:

$$\Delta u(q \mid B, e) = \frac{1}{2} \sum_{v \in \Omega_e} \log \left(1 + \frac{f_q(v)}{f_t(v) + \sum_{b \in B} f_b(v) + \lambda} \right).$$

This discourages selecting several rollout siblings that constrain the same target-local cells from the same directions.

Candidate Tokens And Training Hierarchy

SCONE/Fisher features should enter ARIA-NBV as explicit candidate-token channels:

$$x_{q,e} = \begin{bmatrix} \text{pose features} & \text{target features} & \text{current support} \\ \text{candidate target support} & u_{\text{SCONE}}(q, e) & u_{\text{info}}(q, e) \\ \text{directional novelty} & \text{validity reason} & \end{bmatrix}.$$

The entries are arranged in rows for readability; their concatenation forms one candidate-token vector.

The training/evaluation hierarchy should remain:

1. Cheap actor-visible support features: u_{vis} , u_{info} , and directional novelty.
2. Learned one-step scorer: $\hat{r}_{\psi}^e(s, q)$.
3. Learned finite-horizon value model: $Q_H(s, e, q)$.
4. Oracle evaluation: $J_e^{(H)}$ and endpoint target quality.

SCONE/Fisher features help the model and the data-generation diagnostics. They do not define success.

Concrete Feature Proposal

The minimal implementation implied by the distillation is a `target_support_features` module with three per-candidate features:

$$\phi_{\text{support}}(q_i, e) = [u_{\text{vis}}(q_i, e), u_{\text{info}}(q_i, e), u_{\text{dir}}(q_i, e)].$$

The three components can be:

$$\begin{aligned} u_{\text{vis}} &= \sum_v \hat{p}_{\text{surf}}(v) \hat{p}_{\text{frustum}}(v, q_i), \\ u_{\text{info}} &= \sum_v \log \left(1 + \frac{\hat{p}_{\text{surf}}(v) \hat{p}_{\text{frustum}}(v, q_i)}{n_t(v) + \lambda} \right), \\ u_{\text{dir}} &= \sum_v \hat{p}_{\text{surf}}(v) \hat{p}_{\text{frustum}}(v, q_i) \left[1 - \frac{d_i(v)^\top M_{\text{dir}}(v) d_i(v)}{\text{tr } M_{\text{dir}}(v) + \epsilon} \right]. \end{aligned}$$

This is deliberately much smaller than implementing SCONE or FisherRF literally. It is enough to test whether coverage-support and information-style channels help target-RRI ranking.

Diagnostics And Experiments

Initial experiments should test the auxiliary channels without changing the reward:

check	purpose
Spearman(u_{vis} , target-RRI)	Does actor-visible support correlate with oracle target improvement?
Spearman(u_{info} , target-RRI)	Does diminishing-returns support identify useful target views?
Spearman(u_{dir} , target-RRI)	Does directional novelty help beyond support count?
top-k oracle hit using support features alone	Do auxiliary channels shortlist high-RRI candidates?
per-candidate-family support histograms	Is the bottleneck candidate generation, validity, support, or learned ranking?
selected-family distribution under oracle-1 and oracle-lookahead	Do oracle policies prefer different candidate families than the learned scorer?
support overlap among rollout branches	Are stochastic branches collapsing to duplicates?

The key question is:

Do SCONE/Fisher-inspired actor-visible support features correlate with target-RRI,
and do they improve the learned scorer or Q_H ?

If not, they remain diagnostics. If yes, they become scientifically grounded candidate-token features.

Do Not Adopt

- Do not replace target-RRI with SCONE coverage. Coverage can be high while target geometry quality remains poor.
- Do not require a radiance field, NeRF, or 3DGS state for thesis-core ARIA-NBV experiments.
- Do not make uncertainty reduction the reward unless it is calibrated against target-RRI.
- Do not add SCONE’s full learned occupancy/visibility modules before the finite-candidate target-RRI store, support diagnostics, and rollout/ Q_H contracts are stable.
- Do not interpret poor Q_H results until candidate support, validity masks, reason codes, and candidate-family support strata are audited.

Thesis Wording

SCONE helps ARIA-NBV say: candidate generation is not only pose sampling; it must measure actor-visible target-local support and directional novelty, inspired by volumetric surface-coverage integration.

FisherRF helps ARIA-NBV say: novelty should not be a pure heuristic; candidates that constrain already well-supported target cells are less useful than candidates that constrain uncertain or under-observed target cells.

Together, they support a finite-candidate value model whose candidate token contains pose, target relation, validity, RRI baseline score, SCONE-like visibility support, Fisher-like information gain, and $\mathbb{S}^2$ directional novelty. The final claim is still evaluated by oracle target-RRI and endpoint target reconstruction quality.

Terminology

Geometry

OBB – Oriented Bounding Box: Box described by a center, orientation, and three extents. It approximates an object extent and defines a local target frame. Semantic class and confidence can accompany the detection but are not geometric properties of the box itself.

Metrics

RRI – Relative Reconstruction Improvement: Reduction in reconstruction error relative to the error before one acquisition. Target-specific RRI measures this change on the selected object. Oracle RRI obtains the measurement from privileged reference geometry. Unlike the root-gain reward, its denominator changes with the current reconstruction.

Bibliography

- [1] A. Guédon, P. Monasse, and V. Lepetit, “SCONE: Surface Coverage Optimization in Unknown Environments by Volumetric Integration,” in *Advances in Neural Information Processing Systems*, 2022. [Online]. Available: <https://arxiv.org/abs/2208.10449>
- [2] W. Jiang, B. Lei, and K. Daniilidis, “FisherRF: Active View Selection and Uncertainty Quantification for Radiance Fields using Fisher Information.” [Online]. Available: <https://arxiv.org/abs/2311.17874>
- [3] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis,” in *Computer Vision - ECCV 2020*, 2020, pp. 405–421. doi: 10.1007/978-3-030-58452-8_24.
- [4] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis, “3D Gaussian Splatting for Real-Time Radiance Field Rendering,” *ACM Transactions on Graphics*, vol. 42, no. 4, 2023, doi: 10.1145/3592433.
- [5] e3nn contributors, “e3nn spherical harmonics documentation.” [Online]. Available: https://docs.e3nn.org/en/stable/api/o3/o3_sh.html