POLICY-DEPENDENT VIEW ACQUISITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Selecting camera frames before transmission requires a robot to estimate the value of observations it has not yet received. That value depends on both the views and the policy that consumes them. We formulate conditional view-set value and distinguish decision-level interaction from excess interaction induced by a policy. Selective substitutability training (SST) penalizes excess interaction together with subset regret and action-uncertainty mismatch; a residual coalition scheduler (RCS) selects fresh views from low-rate camera sketches under a per-decision byte cap. Across eight MuJoCo/robosuite manipulation tasks and three training seeds, SST+RCS achieves 69.58% episode success at a 55% communication cap, compared with 62.92% for all-subset training with a Set Transformer. The paired difference is 6.67 percentage points, with a 95% hierarchical bootstrap interval of $[2.78, 10.56]$. A separate eight-seed evaluation shows a similar intermediate-budget gain. An expanded physical evaluation yields a $+7.81$-point paired difference across 576 trials per method. In the main frontier, SST+RCS loses at the smallest cap, ties at full view and has slightly higher mean latency than the reference.

1 INTRODUCTION

A robot approaching an occluded object may need a side view to resolve depth, while an earlier motion can be executed from a wrist camera alone. Active-perception systems address such dependencies by controlling where a sensor looks (Bajcsy, 1988; Cheng et al., 2018). A limited uplink creates a different decision: which fixed-camera full-resolution frames should be requested before they are received? Always sending all views avoids this choice but pays the communication cost at every visual update. Ranking cameras independently is cheaper, yet can miss a pair whose value emerges only when the two views are used together.

The value of a view is also a property of the controller. A well-trained policy may use one image effectively, whereas another policy may need a second image to resolve a representational failure. Sensory-information studies already show that observation loss interacts with policy proficiency and task conditions (Krishna et al., 2025). As Figure 1 illustrates, two views may genuinely resolve an ambiguous decision, or merely compensate for a policy that fails to use information available in one view. A scheduler fitted to the latter can compensate for a policy limitation without removing it.

We describe this distinction using set-valued decision utility. The marginal value of refreshing camera i depends on which other cameras are refreshed. A second difference measures local complementarity or redundancy. Applying the same difference to a decision reference and to a frozen policy gives a signed interaction gap; its positive part is the excess penalized in training. The gap depends on how subset regret varies across subsets, not on its level for one subset. The comparison requires the same utility, conditioning information and continuation protocol.

Selective substitutability training (SST) uses this comparison as a regularizer. It combines control supervision with subset regret, a positive excess-interaction penalty and action calibration. Its purpose is to reduce avoidable reliance on joint observations while retaining the control value of informative combinations. A residual coalition scheduler (RCS) predicts feasible-set values from current low-rate sketches and prior history. Sketches are charged to the communication budget, so the scheduler's decision is based on information that can actually be available before full-frame acquisition.

The strongest closed-loop gain occurs at intermediate communication caps. SST+RCS loses to the reference at the smallest cap and ties it at full view, while the physical-robot difference remains uncertain with 80 trials per method. SST has lower mean excess and subset regret than the other

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

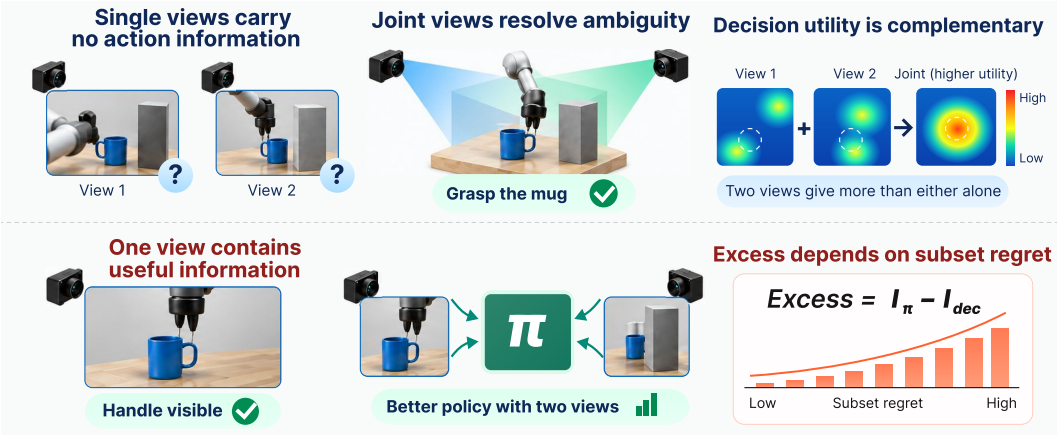

Figure 1: Decision-level and policy-dependent view interaction in a conceptual mug grasp. Top: individually ambiguous views jointly resolve the grasp, so decision utility is complementary. Bottom: the handle is visible in one view, but this policy improves only with a second view, so $I_\pi > I_{\mathrm{dec}}$. The bottom-right panel shows the signed gap $I_\pi - I_{\mathrm{dec}}$, whose positive part is the excess E_π. Its bars are schematic: by Equation (6), the gap follows how subset regret varies across subsets, not its level.

shaping conditions, but its greedy-selection diagnostic remains unfavorable. These results keep policy quality, scheduler quality and deployment cost separate.

The contributions are a common-utility formulation of policy-dependent view interaction; a training and acquisition procedure that uses the formulation under metered observations; and empirical comparisons that separate selection quality, closed-loop success and deployment cost. Paired gains at intermediate budgets are supported by task-level and acquisition diagnostics. Low-budget, full-view and shift results bound the claim.

2 Related Work

Decision Value of Information. Decision theory ranks experiments by enabled decisions and values information through expected utility (Blackwell, 1953; Howard, 1966). Complementarity and substitutability describe interactions among signals (Börgers et al., 2013; Chen & Waggoner, 2017). In robotics, Krishna et al. (2025) relate withheld observations to task and policy properties. We compare camera coalitions for one utility and controller: selected views refresh before the next action, while the others remain stale. This value is policy- and protocol-specific.

Active Perception and View Control. Active perception couples sensing and control (Bajcsy, 1988; Aloimonos et al., 1988; Chen et al., 2011). Learned agents select views for recognition (Johns et al., 2016), reveal occluded objects (Yang et al., 2019), or coordinate camera motion and manipulation (Rasolzadeh et al., 2010; Cheng et al., 2018; Zhang et al., 2023; Wang et al., 2025; Xiong et al., 2025). Those systems move an embodied sensor. We select fixed-camera streams after low-rate sketches under a byte budget, leaving unselected streams stale.

Cost-Aware Observation Selection. Classical submodular maximization, sensor placement and adaptive submodularity justify greedy procedures under diminishing-return and monotonicity assumptions (Nemhauser et al., 1978; Krause et al., 2008b;a; Golovin & Krause, 2011). Submodular active perception and subset selection build on related structure (Satsangi et al., 2018; Das & Kempe, 2011). Our values can be non-monotone, costs are heterogeneous, and the reported ratio averages local minima. Greedy selection is therefore an empirical comparator here, without a cardinality-constrained guarantee. The identity in Section 3 likewise does not imply that lower average excess improves a worst-case set function.

Set-Valued Fusion and Missing Observations. Deep Sets and Set Transformers provide coalition-value baselines (Zaheer et al., 2017; Lee et al., 2019). Missing-input robustness motivates modality dropout and fusion studies (Neverova et al., 2016; Ma et al., 2022); we distinguish missing current

frames from charged refreshes and retained stale frames. We assess calibration separately from accuracy (Guo et al., 2017). The controller uses diffusion-policy action chunks (Chi et al., 2023); our contribution concerns its observation subsets.

3 Conditional View Value

3.1 Acquisition before Transmission

Let $C = \{1, \dots, M\}$ index a fixed camera rig. At visual decision t, history h_t^- contains the instruction, proprioception, controller and action-buffer state, preceding actions, stored images with their ages and link information. A scheduler chooses a fresh-view set $S \subseteq C$. The policy receives the current images from S and stored images from $C \setminus S$; refreshing a view does not change the input schema. Under sketch-based acquisition, the router first receives sketches z_t from every available camera, giving information $h_t = (h_t^-, z_t)$. The feasible family is

$$\mathcal{F}(B_t) = \left\{ S \subseteq C : c_{\text{sketch}} + \sum_{i \in S} c_i \leq B_t \right\}. \tag{1}$$

Costs enter through this constraint, not as a second penalty on the reported task success. The camera-side encoder knows each JPEG's length before transmission and includes those lengths in the charged sketch packet; the router therefore evaluates the feasible family with current encoded payload sizes. The cap covers application payload bytes, including this metadata, but not transport framing. History-only acquisition (P0) does not receive current sketches. Sketch-based acquisition (P1) pays their cost. A privileged comparator (P2) receives all current full frames before selecting; its information advantage is not a deployable communication saving.

Let $V^{\pi}_{\text{post}}(S \mid x_t, h_t, \rho)$ be the expected utility from a subset intervention at simulator state x_t and fixed available history h_t, using a specified horizon and continuation protocol ρ. The online acquisition target averages this value over the latent state conditional on the router's information:

$$Q^{\pi}_{\text{acq}}(S \mid h_t) = \mathbb{E}[V^{\pi}_{\text{post}}(S \mid X_t, h_t, \rho) \mid h_t], \qquad S_t \in \arg\max_{S \in \mathcal{F}(B_t)} Q^{\pi}_{\text{acq}}(S \mid h_t). \tag{2}$$

A realized rollout label and its conditional prediction are different objects. The former can use privileged state offline; the latter cannot use an unrequested full frame online. Executing the scheduler also changes later histories, so offline selection regret and episode success are evaluated separately.

3.2 Interaction and Regret

Fix a history, observation channel and decision utility $U_\rho(X, a)$ that includes a common continuation. The decision reference and policy value are

$$V_{\text{dec}}(S) = \mathbb{E}_{O_S}\left[\max_a \mathbb{E}[U_\rho(X, a) \mid h, O_S]\right], \tag{3}$$

$$V_\pi(S) = \mathbb{E}_{X, O_S, a \sim \pi(\cdot \mid h, O_S)}[U_\rho(X, a)]. \tag{4}$$

Expectations condition on the same prior history. The reference maximizes over actions measurable from the available observations; it does not reveal the hidden state to the deployed policy. For distinct $i, j \notin S$, define

$$I_f(i, j \mid S) = f(S \cup \{i, j\}) - f(S \cup \{i\}) - f(S \cup \{j\}) + f(S). \tag{5}$$

Positive interaction denotes a gain from joint availability beyond the two separate gains. It does not imply that either singleton is useful. Negative interaction denotes local diminishing returns, while a negative first difference denotes non-monotonicity of the evaluated decision rule.

Subset regret is $G_\pi(S) = V_{\text{dec}}(S) - V_\pi(S)$, and positive excess is $E_\pi(i, j \mid S) = [I_\pi(i, j \mid S) - I_{\text{dec}}(i, j \mid S)]_+$. With an exact reference and common utility, $G_\pi \geq 0$. Substitution gives

$$I_\pi - I_{\text{dec}} = -G_\pi(S \cup \{i, j\}) + G_\pi(S \cup \{i\}) + G_\pi(S \cup \{j\}) - G_\pi(S). \tag{6}$$

Thus excess reflects variation in shortfall across subsets. A constant shortfall cancels, and small excess alone does not imply a competent policy. Full-view regret can remain positive. Also, the mean

of the positive part is not the positive part of the mean difference; the two operations are kept separate in every aggregate.

Approximate references replace V_{dec} in measurement. Their agreement cannot establish the exact inequality $G_\pi \geq 0$ or isolate training from architecture as the source of excess. Appendix C gives the algebraic identities and error propagation, and Appendix E evaluates reference agreement without discarding disagreements.

4 Training and Acquisition

4.1 Subset-Compatible Control

The policy encodes each camera image with shared visual weights and includes camera identity, calibration information and frame age in its tokens. Fusion also receives proprioception and the instruction. A camera that is not refreshed contributes its stored frame and its updated age. This makes every measured subset a legal policy input and distinguishes acquisition from replacing an image with an arbitrary mask. The action head predicts a chunk of eight low-level actions; visual-decision latency and the 20 Hz low-level control rate therefore refer to different operations.

Training samples observation subsets alongside the full-view input. At a diagnostic decision point, simulator forks retain the same state, policy checkpoint and random-number streams while varying the fresh-view set. The horizon and subsequent sensing protocol are held fixed within the comparison. For the four-camera rig, all 16 subsets can be represented, including the empty fresh set. The empty set still has history and stale observations and does not mean blind control. The offline-to-online flow is summarized in Figure 2. Restore failures are a separate diagnostic: the recorded audit contains 9 failures in 2400 tested forks, rather than evidence of perfect replay.

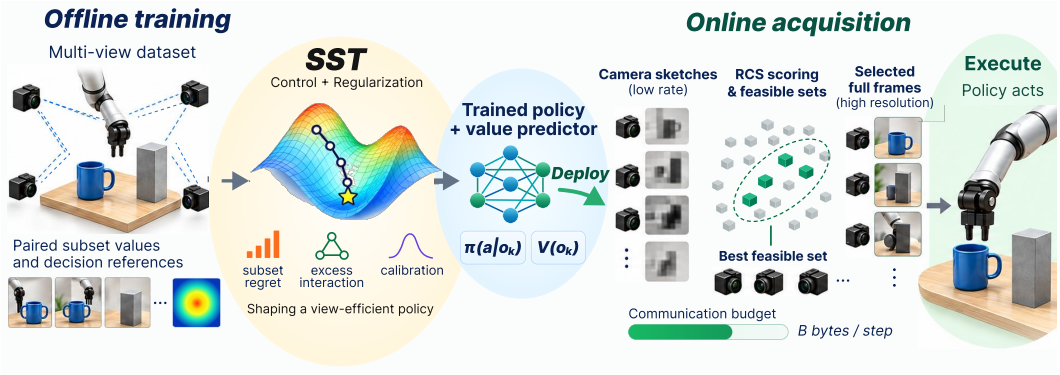

Figure 2: Offline training and online acquisition. Multi-view demonstrations supply SST's control term; paired subset values and decision references supply its subset-regret, excess-interaction and calibration terms. Subset values of the trained policy then fit the RCS value predictor. Online, sketches from every camera arrive first, RCS scores the byte-feasible sets, and the selected full frames refresh the policy input. Labels write a fresh-view set as K (S in the text); $V(o_K)$ is the RCS score $\widehat{Q}(S, z)$ of Equation (9), computed before full frames arrive. Sketches and frames share the per-decision byte budget B_t (Equation (1)).

4.2 Selective Substitutability Objective

Two approximate decision references are used: a privileged critic and a posterior-based action evaluator. Both target the same subset decision utility. The posterior evaluator averages candidate-action utilities over retrieved latent states and then maximizes the average. Maximizing separately for each hidden state would provide extra information and is not the same decision reference. Reference errors, including shared bias, remain relevant even when the two estimators agree in sign.

SST minimizes a control loss together with three regularizers:

$$\mathcal{L} = \mathcal{L}_{\mathrm{ctrl}} + \alpha\mathcal{L}_{\mathrm{sub}} + \beta\mathcal{L}_{\mathrm{exc}} + \eta\mathcal{L}_{\mathrm{cal}}, \qquad \mathcal{L}_{\mathrm{sub}} = \mathbb{E}_S w_S [\widehat{V}_{\mathrm{dec}}(S) - \widehat{V}_\pi(S)]_+. \tag{7}$$

The control term supplies demonstration supervision under the sampled subsets. The subset term compares a policy with the reference for that same subset. The excess term is

$$\mathcal{L}_{\text{exc}} = \mathbb{E}_{S,i,j} w_{S,i,j} \left[\widehat{I}_{\pi}(i,j \mid S) - \widehat{I}_{\text{dec}}(i,j \mid S) \right]_{+}^{2}. \tag{8}$$

For each sampled triple, the policy is evaluated on the base set, both singleton extensions and their joint extension. A differentiable critic supplies action-value gradients; frozen reference targets receive none. The calibration term matches soft-mode entropy and maximum mass to a posterior reference, with $(\alpha, \beta, \eta) = (1, 0.5, 0.2)$. Interaction weights combine subset-size balancing with oracle sign agreement during training; geometry statistics retain all evaluation records. Appendix A specifies the acquisition transaction, posterior values and action-confidence interpretation.

The positive part penalizes interaction above the reference, including cases where both interactions are negative. A positive intrinsic interaction does not automatically switch the penalty off: the term is active whenever the policy interaction exceeds it. The objective is consequently a preference over a regret profile, not a hard constraint preserving every useful synergy. Competing losses, approximation error and optimization can prevent simultaneous improvements. We report excess, subset regret and calibration separately to test that distinction.

4.3 Residual Coalition Scheduling

After policy training, a scheduler predicts set value from the permitted router input z, including sketches under P1. RCS separates singleton and pairwise terms from a residual set term:

$$\widehat{Q}(S, z) = b(z) + \sum_{i \in S} u_i(z) + \sum_{\substack{i<j \\ i,j \in S}} v_{ij}(z) + r(S, z). \tag{9}$$

Camera identity accompanies each sketch, and the set representation does not depend on token ordering. Outcome regression fits the subset labels; a listwise ranking objective emphasizes the feasible candidates, while a regret-weighted term emphasizes costly selection errors. Pairwise terms permit the model to represent conditional value that singleton scores cannot express. The residual accommodates higher-order structure rather than assuming it vanishes after training.

At inference, the scheduler forms the feasible family using the current byte cap and selects its highest-scoring set. The four-camera experiment permits direct enumeration. Selection quality is evaluated on ten feasible candidates at the 55% cap, including the empty set. All selection comparators, including RCS-unshaped, use the same frozen SST policy and candidate values. The separate crossing of all-subset versus SST policies and Set Transformer versus RCS routers compares complete systems.

If predicted values uniformly approximate a fixed acquisition objective within ϵ, and search loses at most δ in predicted value, selection regret is at most $2\epsilon + \delta$. This elementary bound concerns value error on the feasible family. It neither identifies ϵ from mean excess nor guarantees accurate closed-loop control.

5 Experiments

5.1 Evaluation Setup

The simulation evaluation contains table wipe, block stacking, peg insertion, connector plug, shelf grasp, bin retrieval, tool handover and cloth handover. These form static, insertion, occlusion and handover families in a MuJoCo/robosuite setting (Zhu et al., 2020). We use 400 demonstrations per task, $120{,}000$ updates per training run and training seeds $0, 1, 2$. Each frontier point contains eight tasks, three independently trained checkpoints and 30 evaluation episodes per task and seed, giving 720 episode outcomes. Equal task counts make pooled success the task-macro average. Appendix B details the evaluation and cost protocols.

The four-camera nominal frame costs are $26, 26, 34, 34$ kB. P1 sketches average 5.2 kB per visual decision. Labels of $30, 40, 55, 80, 100\%$ correspond to $36, 48, 66, 96, 120$ kB relative to the nominal full-view payload. Budgeted selection uses current encoded lengths; the full-view endpoint bypasses sketches and is an unconstrained comparator, since a nominal 120 kB average does not certify a



Figure 3: Policy geometry. (a) Mean decision interaction, policy interaction and positive excess for the all-subset policy, across four task families and pooled. (b) The fraction of pair/base-set interactions with positive excess for that policy. (c) The reported excess-regret correlation for each policy over 384 decision points. (d) Sign-flip frequency, within- and cross-architecture Jaccard agreement, Kendall correlation and residual variance fraction; percentages are converted to fractions. Panel (b) uses 9216 pooled interaction records, which are not independent decision points. These summaries are shown as point estimates.

120 kB per-decision maximum. The frontier includes a two-policy by two-router comparison and two additional baselines; all-subset training plus Set Transformer is the principal reference. Paired-difference intervals use a hierarchical bootstrap on matched episode records with tasks fixed, whole training-seed IDs resampled across tasks and paired episodes resampled within task-seed cells. Individual task intervals are descriptive Wilson intervals on 90 evaluated episodes and do not account for training-seed dependence. No null-hypothesis test is used.

5.2 Policy Geometry

The all-subset policy has mean decision interaction -0.02573 and mean policy interaction -0.02500, while mean positive excess is 0.01049 (Figure 3a). Average interaction is negative in every family. Positive excess nevertheless occurs in 51.71% of the pair/base-set records because rectification precedes averaging. These statistics describe local departures from the decision reference without establishing that the task families are predominantly complementary. The family fractions occupy a narrow range, 51.04-52.13%.

Policy dependence is measurable but heterogeneous. The matched-quality sign-flip frequency is 9.43%, and critical-subset Jaccard agreement decreases from 0.490 within architecture to 0.322 across architectures (Figure 3d). The reported residual variance fraction after conditioning on full-view performance is 0.9783. These summaries do not identify a causal source of the variation. Excess-regret correlation is 0.149 for all-subset training and 0.508 for SST (Figure 3c); a larger within-policy correlation does not imply that the policy has lower average regret.

SST reduces positive excess relative to dropout from 0.01438 to 0.00940, and subset regret from 0.12776 to 0.11150 (Figure 4a). All-subset training reaches 0.01049 excess and 0.11757 regret, while distillation reaches 0.01004 and 0.11516. SST also has lower ECE, 0.04850, than distillation's 0.05060 and all-subset training's 0.09917. Penalizing all interaction retains the largest mean subset

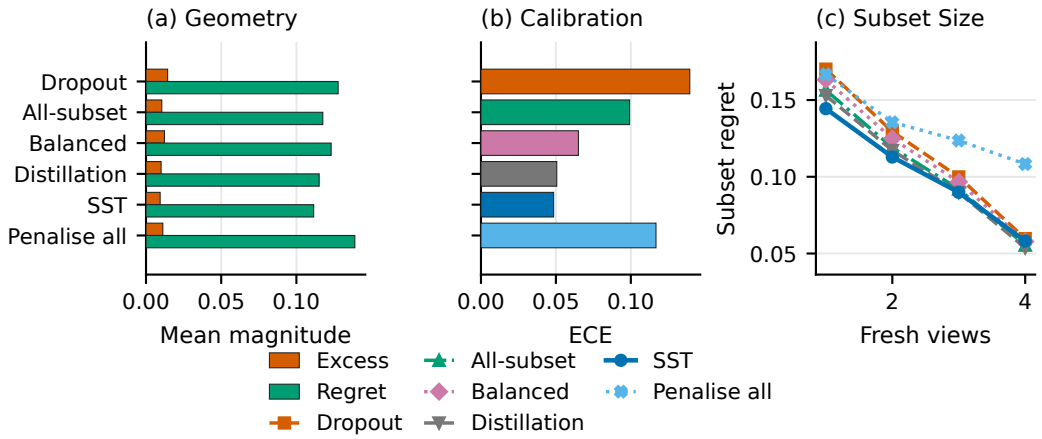

Figure 4: Training-condition comparison. (a) Mean positive excess and subset regret over 384 decision points per condition. (b) Action calibration ECE on 960 ambiguity-state events. (c) Subset regret grouped by the number of fresh views, averaging the subsets at each size. The six policies are dropout, all-subset training, balanced fusion, distillation, SST and penalizing all interaction. Values have no attached uncertainty estimates. Full-view regret is not constrained to zero.

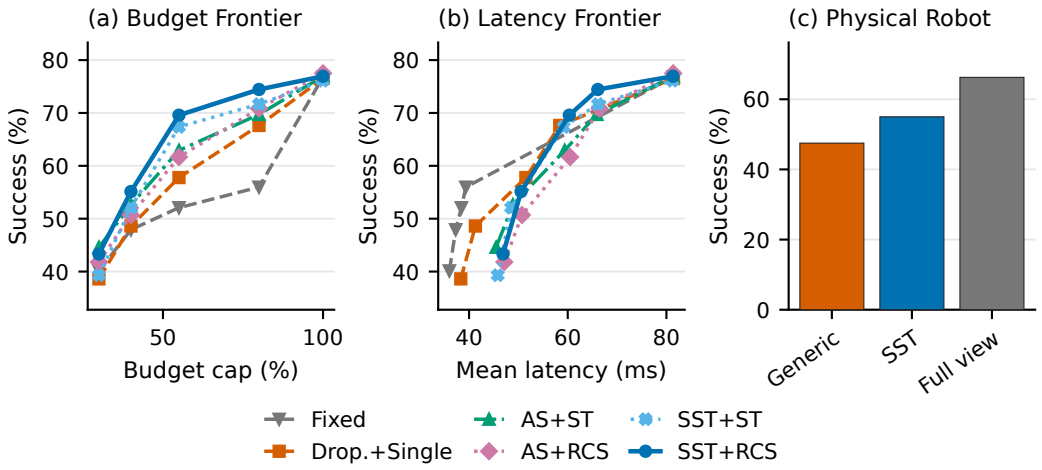

Figure 5: Control and deployment cost. (a) Simulation success at five communication caps, with 720 episodes per method and cap. (b) The same condition-level success summaries against mean end-to-end latency from 1200 visual decisions per condition; this is not an episode-level latency regression. (c) Physical-robot success on four tasks, 20 trials per task: Generic and SST use a 55% cap, while Full view uses 100%. Curves and bars show point estimates; paired uncertainty for the principal comparison is reported in the text. AS denotes all-subset training and ST denotes Set Transformer.

regret, 0.13886. These geometry and calibration measurements are analyzed separately from closed-loop success.

5.3 Budgeted Control

At the 55% cap, SST+RCS achieves 501/720 successes (69.58%), compared with 453/720 (62.92%) for all-subset+Set Transformer. The paired difference is 6.67 percentage points with interval $[2.78, 10.56]$. At 80% the difference is 4.72 points, with interval $[0.97, 8.47]$. At 30%, SST+RCS remains below the reference, 43.33% versus 44.58%; at 40% it reaches 55.14% versus 52.64%; and

at full view both obtain 76.94% (Figure 5a). The strongest separation occurs where several camera pairs first become affordable.

The policy-router crossing at 55% is non-additive. All-subset+RCS obtains 61.67%, while SST+Set Transformer reaches 67.36% and SST+RCS reaches 69.58%. The RCS-labeled contrast is -1.25 points under all-subset training and $+2.22$ under SST, a descriptive interaction of 3.47 points. Within-policy router comparisons hold the controller checkpoint fixed; the two policy-training arms use separate checkpoints. The four-arm interaction is interpreted descriptively. Appendix D gives the complete frontier and paired intervals.

SST+RCS uses 52.17 kB on average at the 66 kB cap, versus 53.98 kB for the reference. Its mean latency is slightly larger, 60.32 versus 59.38 ms, and its 95th-percentile latency is 70.26 versus 68.84 ms. Lower bytes therefore do not imply lower total latency. In the physical comparison, SST reaches $44/80$ successes against $38/80$ for the budgeted generic method and $53/80$ for full view (Figure 5c). The six-trial difference across four tasks remains uncertain at this sample size.

The 55% gain is uneven across tasks. Relative to the reference's 90 episodes per task, SST+RCS adds 12 successes on peg insertion and connector plug, 13 on shelf grasp, 6 on block stacking, 5 on bin retrieval and 4 on tool handover. It loses one on table wipe and three on cloth handover. These differences sum to 48 additional successes over 720 episodes. The two task-level reversals remain visible in Figure 7, which shows all five caps.

5.4 Acquisition and Transfer

RCS+SST selects the best feasible set on 84.11% of diagnostic points, compared with 76.56% for the Set Transformer, 69.27% for Deep Sets and 17.97% for singleton scoring (Figure 6a). Its selection regret is 0.00348 against 0.00590 for the Set Transformer. The whole-ranking Spearman correlation is 0.9172 versus 0.9129; choosing the best set and ranking every candidate remain distinct objectives. The privileged exhaustive reference reaches 100% by directly accessing candidate values.

In the separate online-protocol comparison, P0 history-only acquisition obtains 59.17% success and 0.07122 regret, while P1 sketch+SST reaches 69.58% and 0.00348. The privileged P2 reaches 72.50% and 0.00097, leaving a smaller but visible gap (Figure 6b). P1 sketch+generic uses the same SST control policy and reaches 67.36%; it is distinct from the all-subset reference in the main frontier.

The component study gives the full SST+RCS configuration a 4.5-point difference from the generic reference. Without calibration the margin is 3.2 points, without the oracle gate 2.6, without the pairwise residual 1.8, and with indiscriminate interaction penalization -2.0 (Figure 6c). These separate success-probability summaries are not the counted frontier's 6.67-point effect. Transfer remains bounded: the SST margin is $+3$ points with one camera lost, $+1$ with two lost, -1 under a layout shift and $+1$ for a new checkpoint (Figure 6d).

5.5 Interpreting The Budget Frontier

Under the nominal cost profile, the byte constraint changes which camera identities can be acquired, not just their number. At the 30% cap, a 26 kB frame with sketches costs 31.2 kB and fits, whereas a 34 kB frame costs 39.2 kB and does not. At 40%, every singleton is available but no pair fits. At 55%, five of the six pairs are feasible: the two 34 kB views together exceed the cap once sketches are included. The ten candidates used by the acquisition diagnostic are therefore the empty set, four singletons and five pairs. At 80%, only triples containing both 26 kB views fit; online masks instead use actual payload bytes. This discrete expansion helps explain why a budget frontier cannot be read as a smooth response to additional interchangeable cameras. A change in success can reflect access to a particular view combination as well as a change in selection quality.

The full-view endpoint bypasses sketches and transmits all frames. Router labels no longer specify different acquisition choices there, so the endpoint compares the reported system evaluations rather than estimating a routing effect. Within a constrained condition, mean transmitted bytes measure utilization, while feasibility is enforced for each selected set; the 52.17 kB mean at the 66 kB cap does not imply that a smaller cap would preserve the same decisions. Neither average traffic nor a camera-count limit substitutes for the per-decision byte constraint.

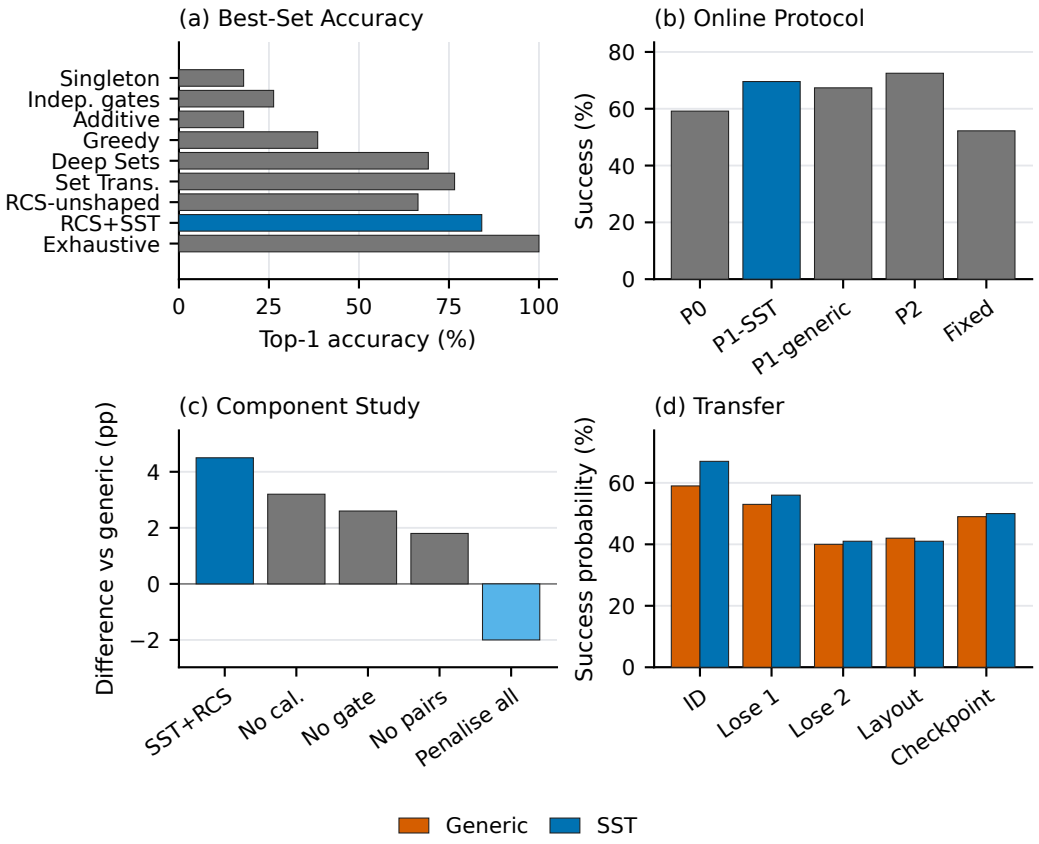

Figure 6: Acquisition and scope. (a) Best-set accuracy over 384 decision points and ten feasible candidates at the 55% cap; all comparators use the same frozen SST control policy. (b) Online-protocol success, 720 episodes per condition. (c) Component-study differences from the generic reference in percentage points. (d) Success-probability summaries under camera loss, layout change and a new checkpoint. Panels (c,d) each use 720 episodes per condition; rounded summaries have no reported intervals and are separate from Figure 5.

6 Conclusion and Limitations

Conditional view value separates an observation channel's decision utility from what a trained policy extracts. SST shapes subset behavior, while RCS selects views before transmission under a byte cap. Across eight simulation tasks and three training seeds, SST+RCS improves paired episode success at the 55% and 80% caps; the 55% difference is 6.67 points with a 95% interval of $[2.78, 10.56]$. A separate eight-seed evaluation yields $+7.19$ points at 55%, and an expanded physical study yields $+7.81$ points across 576 trials per method (Appendix H). These evaluations are analyzed separately. The main frontier retains a loss at the smallest cap, a full-view tie, a layout-shift reversal and slightly higher latency. The smaller $44/80$ versus $38/80$ physical comparison remains uncertain on its own.

Reproducibility Statement

Appendices A and B specify policy inputs, training, acquisition, task success predicates, byte accounting and evaluation units. The main frontier evaluates eight tasks, three training seeds and 30 episodes per task and seed at each cap. Paired episode-level records support the task-fixed hierarchical bootstrap described in Appendix B. Appendix H reports a separate eight-seed evaluation, stress conditions and physical study with their own denominators and seed-level intervals. The two evaluation series are analyzed separately.

AI Use Statement

AI tools assisted with literature search, manuscript organization and writing edits, discussion of experimental design, and code for figures and analysis. The authors are responsible for the experiments, numerical results, references, analyses, and final manuscript.

References

John Aloimonos, Isaac Weiss, and Amit Bandyopadhyay. Active vision. *International Journal of Computer Vision*, 1(4):333–356, 1988.

Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988.

David Blackwell. Equivalent comparisons of experiments. *The Annals of Mathematical Statistics*, 24 (2):265–272, 1953.

Tilman Börgers, Angel Hernando-Veciana, and Daniel Krähmer. When are signals complements or substitutes? *Journal of Economic Theory*, 148(1):165–195, 2013.

Shengyong Chen, Youfu Li, and Ngai Ming Kwok. Active vision in robotic systems: A survey of recent developments. *The International Journal of Robotics Research*, 30(11):1343–1377, 2011.

Yiling Chen and Bo Waggoner. Informational substitutes. *arXiv preprint arXiv:1703.08636*, 2017.

Ricson Cheng, Arpit Agarwal, and Katerina Fragkiadaki. Reinforcement learning of active vision for manipulating objects under occlusions. In *Proceedings of the 2nd Conference on Robot Learning*, volume 87 of *Proceedings of Machine Learning Research*, pp. 422–431, 2018.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin CM Burchfiel, and Shuran Song. Diffusion Policy: Visuomotor policy learning via action diffusion. In *Robotics: Science and Systems (RSS)*, 2023.

Abhimanyu Das and David Kempe. Submodular meets spectral: Greedy algorithms for subset selection, sparse approximation and dictionary selection. In *International Conference on Machine Learning (ICML)*, pp. 1057–1064, 2011.

Daniel Golovin and Andreas Krause. Adaptive submodularity: Theory and applications in active learning and stochastic optimization. *Journal of Artificial Intelligence Research*, 42:427–486, 2011.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On calibration of modern neural networks. In *International Conference on Machine Learning (ICML)*, 2017.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Ronald A. Howard. Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, 2(1):22–26, 1966.

Edward Johns, Stefan Leutenegger, and Andrew J. Davison. Pairwise decomposition of image sequences for active multi-view recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3813–3822, 2016.

Andreas Krause, H. Brendan McMahan, Carlos Guestrin, and Anupam Gupta. Robust submodular observation selection. *Journal of Machine Learning Research*, 9:2761–2801, 2008a.

Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008b.

Arjun Krishna, Edward S. Hu, and Dinesh Jayaraman. The value of sensory information to a robot. In *International Conference on Learning Representations (ICLR)*, 2025.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set Transformer: A framework for attention-based permutation-invariant neural networks. In *International Conference on Machine Learning (ICML)*, volume 97, pp. 3744–3753, 2019.

Mengmeng Ma, Jian Ren, Long Zhao, Davide Testuggine, and Xi Peng. Are multimodal transformers robust to missing modality? *arXiv preprint arXiv:2204.05454*, 2022.

George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978.

Natalia Neverova, Christian Wolf, Graham W. Taylor, and Florian Nebout. ModDrop: Adaptive multi-modal gesture recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(8):1692–1706, 2016.

Babak Rasolzadeh, Mårten Björkman, Kai Huebner, and Danica Kragic. An active vision system for detecting, fixating and manipulating objects in the real world. *The International Journal of Robotics Research*, 29(2–3):133–154, 2010.

Yash Satsangi, Shimon Whiteson, Frans A. Oliehoek, and Matthijs T. J. Spaan. Exploiting submodular value functions for scaling up active perception. *Autonomous Robots*, 42(2):209–233, 2018.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations (ICLR)*, 2021.

Guokang Wang, Hang Li, Shuyuan Zhang, Di Guo, Yanhong Liu, and Huaping Liu. Observe then act: Asynchronous active vision-action model for robotic manipulation. *IEEE Robotics and Automation Letters*, 10(4):3422–3429, 2025.

Haoyu Xiong, Xiaomeng Xu, Jimmy Wu, Yifan Hou, Jeannette Bohg, and Shuran Song. Vision in action: Learning active perception from human demonstrations. In *Proceedings of the 9th Conference on Robot Learning*, volume 305 of *Proceedings of Machine Learning Research*, pp. 5450–5463, 2025.

Jianwei Yang, Zhile Ren, Mingze Xu, Xinlei Chen, David J. Crandall, Devi Parikh, and Dhruv Batra. Embodied amodal recognition: Learning to move to perceive objects. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2040–2050, 2019.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep Sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

Xuechao Zhang, Dong Wang, Sun Han, Weichuang Li, Bin Zhao, Zhigang Wang, Xiaoming Duan, Chongrong Fang, Xuelong Li, and Jianping He. Affordance-driven next-best-view planning for robotic grasping. In *Proceedings of the 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pp. 2849–2862, 2023.

Yuke Zhu, Josiah Wong, Ajay Mandlekar, and Roberto Martín-Martín. robosuite: A modular simulation framework and benchmark for robot learning. *arXiv preprint arXiv:2009.12293*, 2020.