Allocating Reference Memory across Subjects and Time

Spencer Butler (NZ), Charles Powell (BH), Ethan Franklin (QA)

Abstract

Reference memory has to serve both the subject evidence available now and the evidence needed later in a clip. A fixed capacity can be allocated in several ways, each with different information consequences. This article proposes a budgeted comparison of reference-memory policies in subject-preserving video. ARGUS motivates the structured-memory setting. No memory policy is implemented or timed. The design separates capacity, selection, update opportunity, and output assessment so that a future advantage can be attributed to a declared allocation policy.

Capacity is not the same as useful coverage

A memory containing many nearly duplicate observations can provide less view coverage than a smaller but more varied collection. A memory that is useful at the beginning of a clip may be poorly matched to a later requested view. These possibilities make capacity an incomplete description of reference access. An evaluation needs to record which observations are retained, which become available over time, and what information the allocation policy uses when deciding among them.

ARGUS motivates structured identity memory for video generation (Meng et al., 2026b). The present article develops a separate policy-comparison design without modifying its implementation or asserting unmeasured allocation behavior. No resource or preservation results are produced. The scientific question is whether a declared allocation rule changes the useful identity evidence available under a fixed budget, and how a study could distinguish that effect from differences in total source information or update opportunity.

Defining a budgeted policy comparison

A proposed experiment would distinguish the full candidate reference pool from the active conditioning memory. Policies could differ in how they select or replace entries, while the candidate pool and capacity remain recorded. If a policy receives additional views unavailable to another, the comparison concerns a broader workflow rather than allocation alone. Matching the active slot count is therefore necessary for some contrasts but insufficient to establish equal information access across the complete system.

Time creates another information boundary. A policy that can inspect future instructions or output states has different access from one operating only on current information. The protocol should record that distinction explicitly. An offline oracle can be useful as a bounded comparison, provided it is not presented as a deployable policy with the same observations as an online method. Update costs and any additional selection computation should also remain part of the evaluated workflow.

Memory contents should be versioned throughout each generation attempt where the interface permits observation. Those records would link reference choices to output intervals and assessment results. A favorable correlation between selected views and successful identity does not establish causation, since the policy may select easier conditions. Controlled policy contrasts under common tasks provide stronger evidence. If the memory path is unobservable, the study should limit its claim to end-to-end behavior rather than inventing an internal explanation.

Evaluating allocation under temporal demands

Outcomes should include recognizability across the requested trajectory, instruction execution, and assessable coverage. A policy that preserves only an easy initial view should not be described as equally successful throughout a later turn or occlusion. The temporal aggregation rule needs to be fixed before final comparison. At the same time, an unscorable frame should not automatically become a proven memory failure. Independent evidence is needed to distinguish a changed subject from a limitation of the assessment instrument.

The study would preserve subject, instruction, reference-pool, and random-sample links. Many frame observations from one trajectory do not establish performance across many subjects. A matched generation budget would prevent one policy's additional retries from becoming an unexplained advantage. If adaptation or recovery is part of the intended system, it can be evaluated directly, with its reference access and effort reported alongside the resulting outputs rather than excluded from the principal comparison.

The final report could present outcomes across several declared capacities rather than selecting one favorable budget. This would reveal whether an advantage depends on unusually abundant references or persists under tighter constraints. Such a profile remains specific to the source pool and tasks studied. It does not establish an optimal universal memory policy. Its value is a reproducible relationship between allocation choices, information availability, and observed preservation under the workloads that the experiment actually represents.

Replacing references under a fixed memory budget

For the proposed comparison, the key observation is memory update during a planned trajectory. A short reference store must choose between a recent frontal view and a less recent profile needed later in a clip. This hypothetical case poses a question about identity continuity when the requested perspective changes. Record fixed initial slots and view-coverage allocation with the same capacity separately. A policy that consults the realized future clip receives information unavailable to an online selector. Separate an offline oracle from deployable access and log which source observations were active at each event.

An allocation test can prespecify a trajectory with a predictable demand for a previously uncommon view. Compare the reference contents at each time before measuring whether the model used the chosen evidence. For an online policy, reveal only the information available by that step. An offline oracle may receive the full plan, but deserves a separate result. Slot replacement also has a processing cost. Record which view was evicted and whether later behavior exposes a loss of evidence from that earlier choice.

An initial table can record the eligibility and observation status of memory update during a planned trajectory, before any performance statistic is calculated. For the memory update during a planned trajectory, keep undetectable faces in the log and state which cases remained eligible for recognition.

Store conditioning views, generated frames and visibility annotations separately for memory update during a planned trajectory. A useful second view of view-coverage allocation with the same capacity gives each independent group equal weight and retains the row-weighted result for comparison. When examining view-coverage allocation with the same capacity, inspect pairs of frames with similar pose before drawing conclusions about a persistent identity error.

A sensitivity display can show where the conclusion about identity continuity when the requested perspective changes changes as assumptions about the eligible observations change. The evaluation of fixed initial slots can distinguish an incorrect subject from a permitted change in expression, pose or

scene. Collect a separate set of memory update during a planned trajectory under documented conditions before claiming that the contrast generalizes.

Analysis and interpretation

One study controls the false discovery rate for a declared family of hypotheses (Benjamini and Hochberg, 1995). One approach provides guidance on inference under dependence within specified groups (Cameron and Miller, 2015). During validation, a family of endpoints should be declared before evaluation so that reported discoveries reflect every comparison eligible to appear in the result. Repetitions from a common generating condition can have correlated errors, making the number of independent groups more informative than the nominal row count.

When diagnostic endpoints are dependent, the correction should reflect that dependence and use a defined family rather than whichever comparisons looked notable. Single-reference conditioning provides a budget-matched comparison for identifying which errors disappear when extra views are made available to the generator. A relevant contribution examines false discovery rate control when tests are dependent (Benjamini and Yekutieli, 2001). During validation, a relevant contribution conditions identity-preserving image synthesis on a single reference (Wang et al., 2024).

A methodological precedent shows bias from reusing validation for selection and error estimation (Varma and Simon, 2006). Related work separates missing-data assumptions from observed-data summaries (Little and Rubin, 2002). For a prospective comparison, a missingness analysis distinguishes failures from observations that were never eligible and makes the assumptions behind complete-case summaries visible to readers. Reusing a validation split for repeated configuration selection can make the final error estimate optimistic; retain a separate assessment for the chosen rule.

A methodological precedent uses auxiliary views to condition identity-preserving video (Wu et al., 2026). In a controlled test, related work allows regression covariance estimates to account for unequal error variances (White, 1980). Unequal residual variance changes uncertainty even under a common mean model, which matters when a few conditions are much noisier than others. A multiple-view conditioning baseline is most informative when the views add distinct pose evidence rather than repeated copies of a similar crop.

A relevant contribution models repeated longitudinal measurements with random effects (Laird and Ware, 1982). For a prospective comparison, an existing method builds a stacked identity representation from input portraits (Li et al., 2023a). A longitudinal formulation can separate baseline differences between units from changes within the same unit as the sequence progresses. A stacked portrait input permits a test of view coverage, provided the evaluator retains an independent portrait and records near-duplicate input views.

Systems and comparators

A paired edit test can separate matching the reference subject from following a permitted change to composition, clothing or surrounding scene. In an evaluation, a relevant contribution separates global facial structure and local detail in identity-conditioned video (Yuan et al., 2024). One published study uses tuning-free identity customization with attention to preserving image editability (Guo et al., 2024). Separate facial structure and texture checks can identify whether a convincing local detail conceals changes in global appearance over time.

A one-reference adapter supplies a baseline for evaluating whether added views help temporal persistence when motion and instructions are held constant. A separable motion module offers a useful distinction

between the visual quality of a still frame and consistency under movement. A relevant contribution adds a motion module to personalized image-generation systems (Guo et al., 2023). In practical terms, earlier research personalizes video through a face adapter using one reference (He et al., 2024).

First-frame conditioning gives a clear initial anchor; later tests can distinguish ordinary appearance change from drift after a break in visibility. Adaptation comparisons can report trainable parameters together with sample count and compute, keeping the data budget distinct from the architectural change. One approach adapts pretrained models through low-rank parameter updates (Hu et al., 2022). In a controlled test, one study uses first-frame information to guide spatial and temporal consistency (Ren et al., 2024).

A methodological precedent introduces adapters to convey spatial or visual conditions to image diffusion (Mou et al., 2023). In a controlled test, image-conditioned video provides a direct persistence test. Keep the supplied view and later output segment as separate observations. An adapter can hold the type of external condition constant when comparing how faithfully different systems honor an edge, depth or pose specification. A methodological precedent investigates image-to-video generation and video training data curation (Blattmann et al., 2023).

An appearance-aware baseline can be compared with motion-only association under the same detections to locate the source of identity changes. For a prospective comparison, one study uses a diffusion transformer for text-conditioned video generation (Yang et al., 2024). Related work uses appearance information in online association (Wojke et al., 2017). Use a text-conditioned video baseline to assess generic motion with the same clip duration and language instructions.

Measurement and error

One study contrasts precision-recall and receiver operating characteristic summaries under class imbalance (Saito and Rehmsmeier, 2015). Report coverage beside error rates. Rejecting difficult cases can improve conditional accuracy while leaving fewer cases with an answer. One approach reports prediction performance jointly with abstention coverage (Geifman and El-Yaniv, 2017). Low prevalence can make a seemingly small false-positive rate operationally costly, so report precision at plausible recall levels and retain class counts.

One published study uses an additive angular margin to train discriminative face embeddings (Deng et al., 2018). One study examines discrepancies between paired measurement procedures (Bland and Altman, 1986). A paired difference plot can reveal systematic offsets and score-dependent disagreement even when the correlation between two procedures is high. Angular discrimination makes a reasonable verification baseline, provided the threshold is evaluated again on profiles, occlusions and synthetic views outside ordinary portrait capture.

Linking evidence to the evaluation

Earlier research adds image prompting to a text-conditioned diffusion model (Ye et al., 2023). During validation, another study provides face observations with substantial variation in pose and age (Cao et al., 2017). Earlier research constructs a full-reference metric with structure and texture components (Ding et al., 2020). Varying the image-conditioning strength gives a practical control for the trade between copying a reference and following the requested transformation. Separate structure and texture responses are informative when a change in surface detail is permissible but a change in overall shape is consequential. Pose and age diversity help construct cross-condition verification pairs, but transfer tests still need outputs from the intended workflow. Related work asks dataset creators to record collection, intended use and limitations (Gebru et al., 2021). One approach constructs a face attribute dataset with demographic balance (Kärkkäinen and Joo, 2019).

Conclusion

Reference-memory evaluation should distinguish source pools, active capacity, update access, and selection effort. The proposed design records those boundaries and connects them to temporal identity and task outcomes. It offers no new allocation result. A future comparison using these records could show when a policy makes better use of a fixed evidence budget, while avoiding an unsupported attribution of end-to-end improvements to memory organization alone.

References

Wu, Mingyang, et al. "ConsID-Gen: View-Consistent and Identity-Preserving Image-to-Video Generation." *arXiv preprint arXiv:2602.10113*, 2026. https://arxiv.org/abs/2602.10113.

Li, Zhen, et al. "PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding." *arXiv preprint arXiv:2312.04461*, 2023. https://arxiv.org/abs/2312.04461.

Ren, Weiming, et al. "ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation." *arXiv preprint arXiv:2402.04324*, 2024. https://arxiv.org/abs/2402.04324.

Blattmann, Andreas, et al. "Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets." *arXiv preprint arXiv:2311.15127*, 2023. https://arxiv.org/abs/2311.15127.

White, Halbert. “A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity.” *Econometrica*, vol. 48, no. 4, 1980, pp. 817-838. https://doi.org/10.2307/1912934.

Laird, Nan M. and James H. Ware. “Random-Effects Models for Longitudinal Data.” *Biometrics*, vol. 38, no. 4, 1982, pp. 963. https://doi.org/10.2307/2529876.

Guo, Yuwei, et al. "AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning." *arXiv preprint arXiv:2307.04725*, 2023. https://arxiv.org/abs/2307.04725.

Hu, Edward J., et al. “LoRA: Low-Rank Adaptation of Large Language Models.” *International Conference on Learning Representations*, 2022. https://openreview.net/forum?id=nZeVKeeFYf9.

Ding, Keyan, et al. "Image Quality Assessment: Unifying Structure and Texture Similarity." *arXiv preprint arXiv:2004.07728*, 2020. https://arxiv.org/abs/2004.07728.

Saito, Takaya and Marc Rehmsmeier. “The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets.” *PLOS ONE*, vol. 10, no. 3, 2015, pp. e0118432. https://doi.org/10.1371/journal.pone.0118432.

Gebru, Timnit, et al.. "Datasheets for datasets" *Communications of the ACM*, 2021. https://doi.org/10.1145/3458723.

Cameron, A. Colin, and Douglas L. Miller. “A Practitioner’s Guide to Cluster-Robust Inference.” *Journal of Human Resources*, vol. 50, no. 2, 2015, pp. 317-372. https://doi.org/10.3368/jhr.50.2.317.

Benjamini, Yoav and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 57, no. 1, 1995, pp. 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Meng, Zijie, et al. "Argus: Stacked multi-view identity mosaic injection for subject-preserving video generation." *arXiv preprint arXiv:2606.11670* (2026).

Mou, Chong, et al. "T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models." *arXiv preprint arXiv:2302.08453*, 2023. https://arxiv.org/abs/2302.08453.

Wang, Qixun, et al. "InstantID: Zero-shot Identity-Preserving Generation in Seconds." *arXiv preprint arXiv:2401.07519*, 2024. https://arxiv.org/abs/2401.07519.

Wojke, Nicolai, et al. "Simple Online and Realtime Tracking with a Deep Association Metric." *arXiv preprint arXiv:1703.07402*, 2017. https://arxiv.org/abs/1703.07402.

Yang, Zhuoyi, et al. "CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer." *arXiv preprint arXiv:2408.06072*, 2024. https://arxiv.org/abs/2408.06072.

Varma, Sudhir and Richard Simon. “Bias in error estimation when using cross-validation for model selection.” *BMC Bioinformatics*, vol. 7, no. 1, 2006. https://doi.org/10.1186/1471-2105-7-91.

Kärkkäinen, Kimmo and Jungseock Joo. "FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age" *arXiv preprint arXiv:1908.04913*, 2019. https://arxiv.org/abs/1908.04913.

Geifman, Yonatan, and Ran El-Yaniv. “Selective Classification for Deep Neural Networks.” *Advances in Neural Information Processing Systems*, vol. 30, 2017. https://arxiv.org/abs/1705.08500.

Ye, Hu, et al.. "IP-Adapter: Text Compatible Image Prompt Adapter for Text-to-Image Diffusion Models" *arXiv preprint arXiv:2308.06721*, 2023. https://arxiv.org/abs/2308.06721.

Yuan, Shenghai, et al. "Identity-Preserving Text-to-Video Generation by Frequency Decomposition." *arXiv preprint arXiv:2411.17440*, 2024. https://arxiv.org/abs/2411.17440.

Bland, J. Martin, and Douglas G. Altman. "Statistical Methods for Assessing Agreement between Two Methods of Clinical Measurement." *The Lancet*, vol. 327, no. 8476, 1986, pp. 307-310. https://doi.org/10.1016/S0140-6736(86)90837-8.

Deng, Jiankang, et al.. "ArcFace: Additive Angular Margin Loss for Deep Face Recognition" *arXiv preprint arXiv:1801.07698*, 2018. https://arxiv.org/abs/1801.07698.

He, Xuanhua, et al. "ID-Animator: Zero-Shot Identity-Preserving Human Video Generation." *arXiv preprint arXiv:2404.15275*, 2024. https://arxiv.org/abs/2404.15275.

Guo, Zinan, et al. "PuLID: Pure and Lightning ID Customization via Contrastive Alignment." *arXiv preprint arXiv:2404.16022*, 2024. https://arxiv.org/abs/2404.16022.

Benjamini, Yoav and Daniel Yekutieli. “The control of the false discovery rate in multiple testing under dependency.” *The Annals of Statistics*, vol. 29, no. 4, 2001, pp. 1165-1188. https://doi.org/10.1214/aos/1013699998.

Cao, Qiong, et al.. "VGGFace2: A dataset for recognising faces across pose and age" *arXiv preprint arXiv:1710.08092*, 2017. https://arxiv.org/abs/1710.08092.

Little, Roderick J. A., and Donald B. Rubin. *Statistical Analysis with Missing Data*. Wiley, 2002. https://doi.org/10.1002/9781119013563.