How Much Alignment Capacity Does EEG-to-Visual Decoding Need? Retrieval is Preserved with 42% Fewer Parameters

Anonymous authors
Paper under double-blind review

Abstract

EEG-to-visual decoding increasingly maps noisy neural responses into expressive pretrained visual spaces that support natural-image retrieval and generation, yet the amount of learned alignment capacity required to preserve useful visual information remains unclear. We isolate this question in ENIGMA by replacing its dense 1024×1024 final residual transformation with rank-224 factors while holding the parent encoder, visual targets, training objective, participant protocol, and evaluation fixed. The intervention reduces trainable EEG-model parameters from 1.405M to 0.815M (41.97%). In a prospectively frozen confirmation across five development-independent participants, mean 200-way single-trial Top-1 is 8.783% for ENIGMA and 8.909% for the compact model, with both one-sided 95% lower bounds above a prespecified -0.5-percentage-point retention margin. Compression also lowers effective rank, preserves ENIGMA-compact representation geometry under centered kernel alignment (CKA) and local image neighborhoods, and slightly increases correct-image cosine alignment. At 80 repetitions, the compact representation satisfies a prespecified DINO-content retention criterion when conditioning a fixed SDXL-Turbo/IP-Adapter generator with shuffled-EEG controls. Thus, substantial residual capacity is removable while preserving confirmed retrieval and high-repeat downstream conditioning, making alignment capacity an explicit design variable for EEG-to-visual decoders.

1 Introduction

EEG visual decoding now poses a strikingly asymmetric learning problem: millisecond-scale, low-signal-to-noise physiological measurements must be placed into visual representation spaces learned from vastly larger image corpora. Large natural-image benchmarks such as THINGS-EEG2 have made that problem measurable at scale (Gifford et al., 2022), while methods such as NICE and DreamDiffusion have shown that an aligned EEG representation can support both zero-shot retrieval and image generation (Song et al., 2024; Bai et al., 2024). More recent systems have strengthened the visual side of the bridge through multi-subject modeling, richer visual teachers, hierarchical targets, subject-aware alignment, and repeat-aware inference (Kneeland et al., 2026; Wang and Chen, 2026; Tang et al., 2026; Jiang et al., 2026; Cui et al., 2026). As the alignment machinery becomes more expressive, a fundamental design question comes into focus: **how much learned capacity is actually required to preserve the visual information recoverable from EEG?**

We study this question inside ENIGMA, where the final residual alignment block contains a dense 1024×1024 learned transformation (Kneeland et al., 2026). This concentration of parameters creates a controlled setting for a capacity intervention. We replace that dense residual update with a rank-224 factorization while holding the parent EEG encoder, visual targets, training objective, participant protocol, and retrieval evaluation fixed. The resulting model falls from **1,405,243 to 815,419 trainable EEG-model parameters**, a **41.97% reduction**. After the architecture, training rule, and statistical criterion were frozen, we confirmed the comparison on **five development-independent participants with two seeds each**. Mean 200-way single-trial Top-1 changes from **8.783%** for the ENIGMA reference to **8.909%** for the compact model, and both prespecified one-sided lower bounds clear the -0.5**-percentage-point** accuracy-retention margin (Fig. 1).

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.

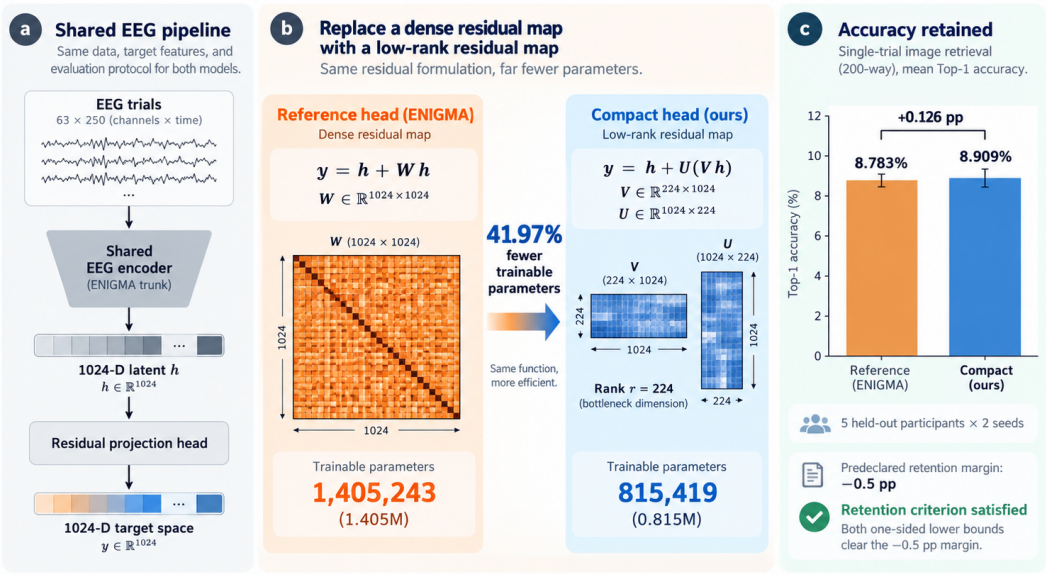

Figure 1: **Rank-constrained alignment removes 41.97% of trainable EEG-model parameters while satisfying prospective single-trial retention.** The final dense 1024×1024 ENIGMA residual transformation is replaced by rank-224 factors while the 1,024-D identity pathway and the rest of the encoder remain fixed. Across five development-independent participants and two seeds, mean Top-1 is 8.783% for ENIGMA and 8.909% for compact; both one-sided 95% lower bounds exceed the prespecified -0.5 pp margin.

The retained accuracy is accompanied by a coherent change in representation structure. Across one, four, and eighty EEG repetitions, the compact code uses fewer effective dimensions while preserving closely matched ENIGMA–compact geometry and nearly unchanged local image-neighborhood structure. Mean cosine alignment with the correct image representation increases slightly at every tested repetition count. At $k = 80$, the same compact EEG representation also satisfies the prespecified downstream DINO-content retention criterion when used to condition a fixed SDXL-Turbo/IP-Adapter generator. The evidence therefore converges on a capacity-allocation result: **a substantially smaller learned residual update can preserve the visual code needed for retrieval and high-repeat generative conditioning.**

Our contributions are threefold:

1. **Controlled capacity reduction with prospective confirmation.** We isolate the final ENIGMA residual transformation, reduce the complete EEG model by **41.97%**, and confirm single-trial accuracy retention across five independent participant-specific fits.
2. **A representation-level account of what compression changes.** Centered kernel alignment (CKA), correct-image alignment, neighborhood preservation, and effective rank show that the compact model concentrates the learned code into fewer active directions while retaining its task-relevant visual organization.
3. **Downstream validation of retained information.** Under a fixed generator, fixed stimuli, paired diffusion noise, and shuffled-EEG controls, the compact representation satisfies the prespecified visual-content retention criterion in the high-repeat setting.

2 EEG-to-Visual Alignment and the Capacity Hypothesis

Visual EEG decoding has shifted from predicting a closed set of labels toward learning representations that can meet images inside pretrained visual spaces. THINGS-EEG2 made this formulation practical at scale by pairing natural-image viewing with repeated EEG measurements across a large visual vocabulary (Gifford et al., 2022). Once EEG is mapped into a visual representation, the

same code can support retrieval, reconstruction, and other downstream uses. That shift places the alignment representation itself at the center of the problem.

2.1 From EEG responses to pretrained visual spaces

NICE established a strong retrieval formulation by contrastively aligning EEG with image representations and evaluating zero-shot recognition over natural images (Song et al., 2024). DreamDiffusion followed the complementary generative path, combining EEG representation learning with CLIP alignment and a pretrained diffusion model to reconstruct viewed content (Bai et al., 2024). Both lines benefit from pretrained visual models such as CLIP, whose image-language space supplies semantic structure that the EEG encoder can learn to approach (Radford et al., 2021).

This formulation changes the modeling problem. The pretrained visual model supplies a large-scale semantic geometry, and the EEG network's task becomes recovering enough neural information to enter that space. Retrieval asks whether that representation lands near the correct image, while generation asks whether it remains informative enough to condition a separate visual decoder.

2.2 Recent progress has strengthened the alignment target

Recent systems have improved EEG-to-visual decoding largely by changing the information presented to the alignment process or the way that information is organized. ENIGMA emphasizes a compact multi-subject architecture, rapid subject adaptation, and parameter-efficient EEG-to-image decoding (Kneeland et al., 2026). CORTIVA combines complementary visual teachers at the candidate-score level (Wang and Chen, 2026); HCF exploits hierarchical structure from intermediate visual representations (Tang et al., 2026); SAMGA introduces subject-aware multi-granularity alignment (Jiang et al., 2026); and NEAR uses high-repeat neural structure to anchor few-repetition retrieval (Cui et al., 2026). A separate scaling analysis across non-invasive modalities shows that decoding performance continues to benefit from more neural data per subject (Banville et al., 2025).

Together, these results identify several productive axes for better decoding: richer visual supervision, hierarchical structure, subject adaptation, repeated-measurement handling, and data scale. This leaves a distinct design axis comparatively underexamined: once the visual target, training exposure, and decoding objective are fixed, how much learned alignment capacity is actually needed to carry the useful EEG information?

2.3 Capacity as a controlled variable

Low-rank factorization has long provided a direct way to expose redundant neural-network parameters (Denton et al., 2014). Representation-similarity methods such as centered kernel alignment provide a complementary way to test whether a smaller model preserves the geometry of the original representation (Kornblith et al., 2019). This make alignment capacity an experimental variable.

We hypothesize that ENIGMA's dense final residual transformation allocates more learned degrees of freedom than are required for EEG-to-visual alignment. The key test therefore holds the parent encoder, visual targets, training objective, participant protocol, and retrieval evaluation fixed while constraining only the learned residual correction. Section 3 implements exactly that intervention.

3 Compact Alignment and Confirmation Protocol

ENIGMA maps EEG into a 1,024-dimensional visual representation and concentrates much of its trainable capacity in the final residual transformation. We intervene only on that block, leaving the parent encoder, visual targets, objective, and evaluation unchanged. This isolates residual capacity as the controlled variable in the comparison.

3.1 Rank-constrained residual correction

After the shared temporal/spatial trunk, ENIGMA flattens four feature maps over 46 positions and projects the resulting 184 values to $u \in \mathbb{R}^{1024}$. Its final deterministic residual mapping is

$$F_{\mathrm{ref}}(u) = \mathrm{LN}(u + W\phi(u) + b)\,, \qquad W \in \mathbb{R}^{1024\times 1024}, \tag{1}$$

where LN denotes LayerNorm, ϕ is GELU, and $b \in \mathbb{R}^{1024}$ is the learned bias. The compact model replaces only W with

$$W = UV, \qquad U \in \mathbb{R}^{1024\times 224}, \quad V \in \mathbb{R}^{224\times 1024}, \tag{2}$$

while retaining the 1,024-dimensional identity pathway.

Theorem 1 (Rank-constrained residual update). For $\phi(u) \in \mathbb{R}^d$ and $\Delta(u) = UV\phi(u)$ with $U \in \mathbb{R}^{d\times r}$ and $V \in \mathbb{R}^{r\times d}$, every residual update lies in $\mathrm{col}(U)$, whose dimension is at most r.

Proof. $\mathrm{rank}(UV) \le \min\{\mathrm{rank}(U), \mathrm{rank}(V)\} \le r$, so $UV\phi(u) \in \mathrm{col}(U)$ for every u. The identity contribution u is carried separately. □

For $d = 1024$ and $r = 224$, the residual matrix falls from 1,048,576 to 458,752 weights. The complete EEG model therefore falls from 1,405,243 to 815,419 trainable parameters, 41.97% less.

Rank 224 was fixed before confirmation as the operating point from a development-stage capacity-allocation design rather than selected by a rank sweep. On S01–S02 (two seeds), the corresponding compact-only arm reached 7.067% single-trial Top-1 versus 6.803% for dense ENIGMA and was carried forward to confirmation. A subsequent development-only sweep over $r \in \{64, 128, 224, 512\}$, together with dense ENIGMA, showed closely grouped low-rank performance and no monotonic retrieval benefit from increasing residual rank (see Appendix B).

3.2 Matched training and retrieval evaluation

We use prepared THINGS-EEG2 arrays at 250 Hz: 63 channels × 250 samples per trial, 16,540 training images with four recordings each, and 200 test images with 80 recordings each. Confirmation uses participants S03–S07, seeds 42 and 43, and separately fitted participant-specific ENIGMA and compact models, for 20 matched fits. All EEG layers train from scratch for 50 epochs with batch size 512 and 6,450 optimizer updates. Both arms receive identical shuffled trial indices and the same fixed pool of 66,048 training trials per epoch. The source-matched objective is raw image-feature MSE plus 0.5 symmetric image contrastive loss and 0.05 symmetric text contrastive loss; optimizer and implementation details are provided in the supplementary material.

The primary endpoint is 200-way single-trial Top-1 retrieval over all 16,000 test trials per participant. Outcomes are averaged within image before seed and participant aggregation. Exact ties receive expected credit under a uniform ordering within the tied rank block, with an independent sorting-based check. Secondary analyses use $k \in \{2, 4, 8, 16, 32, 80\}$ repetitions under fixed waveform-first and representation-first aggregation paths.

3.3 Prospective retention criterion

Let $\Delta = A_{\mathrm{cmp}} - A_{\mathrm{ref}}$ denote the compact-minus-reference single-trial Top-1 difference. The prespecified practical margin is $m = 0.005$ (0.5 percentage points), carried forward from the development plan as a fixed decision gate rather than estimated from confirmation outcomes; Appendix A.6 reports the corresponding margin sensitivity. This gives

$$H_0 : \Delta \le -m, \qquad H_1 : \Delta > -m. \tag{3}$$

Retention requires both a one-sided 95% participant-t lower bound and a one-sided 95% crossed participant/image bootstrap lower bound (10,000 draws) to exceed $-m$, with reference accuracy above 200-way chance in every fit. Superiority is evaluated only after retention, requiring both lower bounds above zero and a one-sided exact participant sign-flip $p \le 0.05$. This frozen design asks one question: whether the rank-constrained head preserves retrieval across independent participants.

4 Main Result: Much of the Alignment Capacity Is Removable

The compact head removes 589,824 trainable weights while preserving the confirmed retrieval behavior. Table 1 reports the primary endpoint and Figure 2 shows its participant + repetition structure.

Table 1: **Controlled confirmation result.** The compact residual head uses 41.97% fewer trainable EEG-model parameters while satisfying the accuracy-retention criterion. Values are means across five development-independent participants and two training seeds. The compact-minus-reference Top-1 difference is +0.126 pp; one-sided 95% lower bounds are -0.106 pp (participant-t) and -0.131 pp (crossed participant/image bootstrap), both above the -0.500 pp retention margin.

Model	Parameters	Top-1 (%)	Top-5 (%)	MRR
ENIGMA reference	1,405,243	8.783	25.777	0.1832
Compact residual head	815,419	8.909	26.077	0.1853

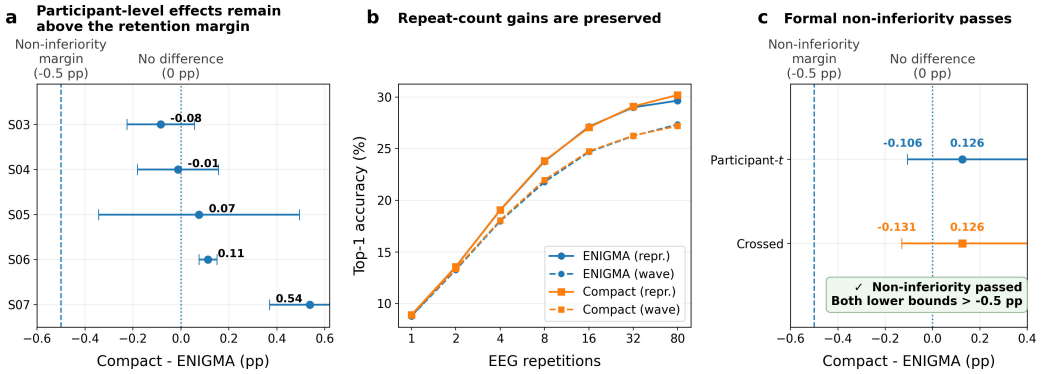

Figure 2: **The compact model preserves the retrieval profile across participants and repetition counts.** (a) Seed-averaged participant effects all lie above the prespecified -0.5 pp retention boundary. (b) ENIGMA and compact follow closely matched Top-1 trajectories under both fixed repeat-aggregation paths. (c) The compact-minus-ENIGMA mean difference is $+0.126$ pp; participant-t and crossed participant/image one-sided 95% lower bounds are -0.106 and -0.131 pp, respectively.

4.1 Prospective single-trial retention

Across five development-independent participants and two seeds, mean 200-way single-trial Top-1 is **8.783%** for ENIGMA and **8.909%** for compact, a difference of **+0.126 pp**. Top-5 changes from 25.777% to 26.077%, and mean reciprocal rank (MRR) from 0.1832 to 0.1853. The participant-t and crossed participant/image one-sided 95% lower bounds are -0.106 and -0.131 **pp**, both above the -0.500 pp retention boundary. The confirmed conclusion is retention under the prespecified margin; the stricter superiority test gives an exact one-sided participant sign-flip $p = 0.1875$.

4.2 Participant replication and repeated EEG

Seed-averaged compact-minus-reference Top-1 effects for S03–S07 are -0.084, -0.012, $+0.075$, $+0.112$, and $+0.537$ pp, respectively; participants are the confirmation units, with seeds and repetitions contributing within-participant precision. Repeated EEG produces closely matched trajectories. At $k = 80$, representation-first Top-1 is 29.65%/30.20% for ENIGMA/compact, while waveform-first Top-1 is 27.35%/27.20%. The retention decision is attached to $k = 1$; the curves show that the reduced-capacity model follows the parent model across increasing neural evidence.

5 What Capacity Was Removed?

Retrieval shows that task-relevant information survives compression; we next examine how the final 1,024-dimensional representation changes. Figure 3 combines global geometry, target alignment, local neighborhoods, effective dimensionality, and a shared-coordinate visualization.

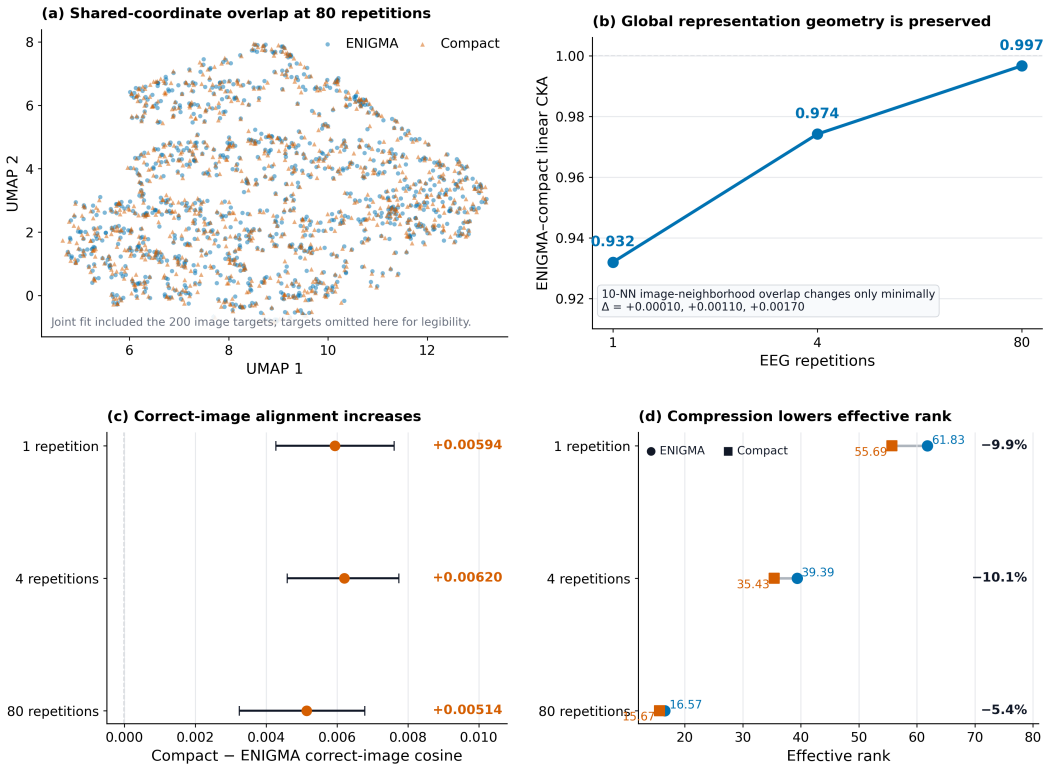

Figure 3: **Compression concentrates the representation while preserving its visual organization.** (a) Shared-coordinate UMAP for the fixed $k = 80$ EEG endpoints; the joint fit also includes image targets. (b) ENIGMA–compact CKA rises from 0.932 to 0.997 across the tested repetition counts. (c) Correct-image cosine alignment is slightly higher for compact at $k = 1, 4, 80$. (d) Effective rank is lower under compression. Original-space CKA, cosine, neighborhood-overlap, and rank measurements carry the quantitative interpretation.

5.1 Global geometry and effective dimensionality

For matched centered representation matrices X and Y, centered linear CKA compares sample geometry through their Gram matrices $K = XX^\top$ and $L = YY^\top$ using the Frobenius inner product and norm:

$$\mathrm{CKA}(X, Y) = \frac{\langle K, L \rangle_F}{\|K\|_F \|L\|_F}. \tag{4}$$

ENIGMA–compact CKA is 0.932, 0.974, and 0.997 at $k = 1, 4, 80$. For singular values s_j of the centered, unit-normalized representation, effective rank is $r_{\mathrm{eff}} = \exp(-\sum_j p_j \log p_j)$ with $p_j = s_j^2 / \sum_\ell s_\ell^2$. It falls from 61.83 to 55.69, 39.39 to 35.43, and 16.57 to 15.67 at the same counts. Compression therefore concentrates the code into fewer active directions while leaving global geometry closely matched.

5.2 Visual alignment and local neighborhoods

Mean cosine similarity to the correct image representation rises from 0.0987 to 0.1047 at $k = 1$, 0.1456 to 0.1518 at $k = 4$, and 0.1892 to 0.1944 at $k = 80$. Compact-minus-ENIGMA gains are positive under both participant-t and crossed participant/image lower bounds at all three counts. Ten-neighbor overlap with image-space neighborhoods remains nearly unchanged (0.0872/0.0873, 0.1238/0.1250, and 0.1565/0.1582 for ENIGMA/compact). The shared UMAP is used only as a qualitative common-coordinate view; the interpretation rests on the original-space metrics. Together, the results are consistent with redundant degrees of freedom in the dense residual correction.

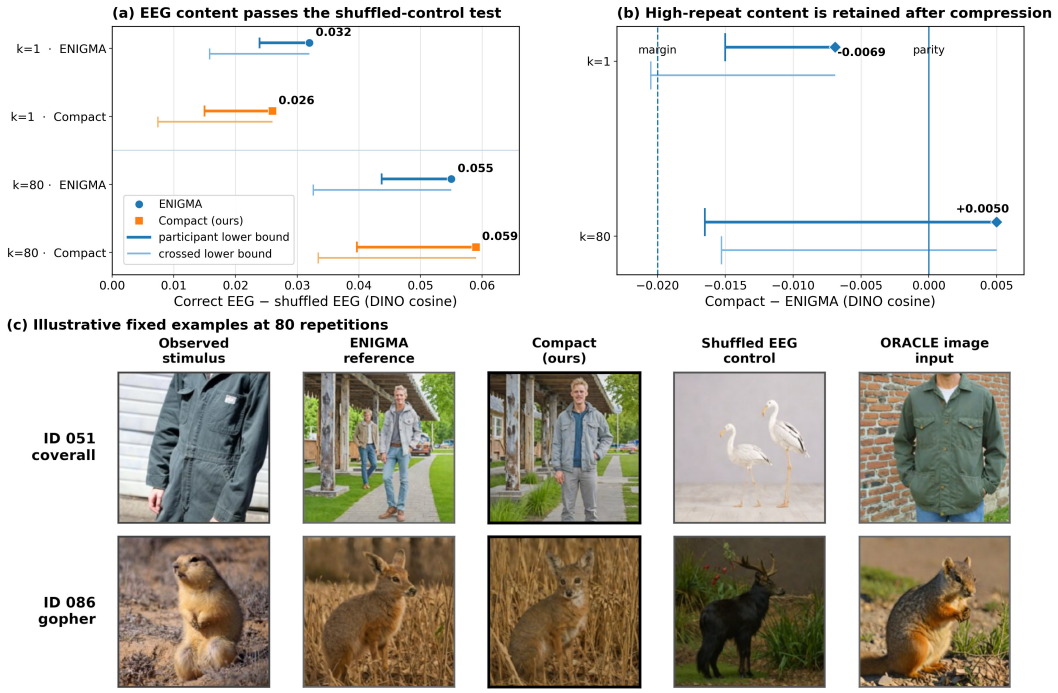

Figure 4: **EEG conditioning carries stimulus-related visual content, and the compact representation retains that content at high repetition.** (a) Correctly conditioned generations exceed shuffled-EEG controls for both ENIGMA and Compact at $k = 1$ and $k = 80$. Points show the mean correct-minus-shuffled DINO cosine; thick and thin intervals show the participant-based and crossed participant/image one-sided 95% lower bounds, respectively. (b) Compact-minus-ENIGMA DINO cosine is -0.0069 at $k = 1$ and $+0.0050$ at $k = 80$. At $k = 80$, both lower bounds remain above the prespecified -0.02 retention margin; at $k = 1$, the crossed lower bound reaches -0.0205 and narrowly misses the margin. (c) Two illustrative examples from the fixed 32-image S03 set at $k = 80$ show the observed stimulus, ENIGMA reference conditioning, Compact conditioning (ours), shuffled-EEG control, and oracle image conditioning under matched diffusion noise. The examples are qualitative illustrations of conditioning behavior.

6 Is the Compressed Representation Useful Downstream?

A stronger test asks whether the compact EEG code remains useful for conditional image generation. We condition a fixed SDXL-Turbo/IP-Adapter generator with saved 1,024-dimensional EEG representations and evaluate generated images with frozen DINOv2-small features. Figure 4 combines fixed examples, shuffled-EEG controls, and the retention analysis (more examples in the appendix).

6.1 Controlled Generation and High-Repeat Retention

For 32 stimuli selected before generation, all five confirmation participants contribute ENIGMA, compact, and shuffled-EEG conditions with paired diffusion noise; encoder seed 42 is fixed. DINO content similarity is the cosine between unit-normalized generated-image and stimulus features. At $k = 80$, correct conditioning exceeds shuffled conditioning by 0.055 for ENIGMA and 0.059 for compact, with positive participant and crossed-image lower bounds. Mean DINO cosine is 0.0769 versus 0.0819, and compact-minus-ENIGMA is $+0.0050$. Both one-sided lower bounds (-0.0165 and -0.0153) clear the -0.02 retention margin. DINO-based 200-way Top-1 is 6.25% vs. 8.125%.

6.2 Single-Trial Boundary

At $k = 1$, correct conditioning still exceeds shuffled conditioning for both models, showing EEG-dependent generated content. Compact-minus-ENIGMA cosine is -0.0069; the participant-t lower

Table 2: **External context under distinct protocols.** The controlled causal comparison in this paper is ENIGMA versus compact ENIGMA. NICE is independently reproduced on the same five participant identities and 200-image test gallery but uses its public CLIP-768 target and mean-repetition training. Published 2026 values retain their method-specific protocols. Parameter counts are trainable EEG-model parameters for the ENIGMA rows and EEG inference parameters for NICE.

Method	Provenance	Protocol / visual supervision	EEG params	Top-1 %
ENIGMA-derived reference	Kneeland et al. (2026); reproduced here	ENIGMA 1024-D target; individual-trial training; $k = 80$ waveform	1,405,243	27.35
Compact ENIGMA	This work	Same target/training; rank-224 residual; $k = 80$ waveform	815,419	27.20
NICE reproduction	Song et al. (2024); reproduced here	Official CLIP-768 target; mean-repetition training	1,802,496	13.25
NICE published	Song et al. (2024)	Published 200-way zero-shot protocol	–	15.60

bound is -0.0150, while the crossed participant/image lower bound is -0.0205, just beyond the same margin. The downstream retention claim therefore belongs to the many-repetition setting, giving a clear boundary for the positive high-SNR result.

7 External Context: NICE and Current EEG Decoders

The controlled ENIGMA comparison isolates capacity. Table 2 adds an independently reproduced same-dataset baseline and recent published systems while keeping their protocol differences explicit.

Using the official NICE construction and CLIP-768 targets, 10 reproduced fits across the same five participant identities and two seeds reach **13.25% Top-1 / 40.20% Top-5** with **1,802,496 EEG inference parameters**; the published ICLR 2024 Top-1 is 15.6% (Song et al., 2024). On the same participants and 200-image identities, compact ENIGMA reaches 27.20%/58.00% at $k = 80$ waveform evaluation with 815,419 parameters. NICE uses mean-repetition training in a CLIP-768 target space, so this is same-dataset external context; the causal architectural comparison remains ENIGMA versus compact ENIGMA.

Recent work shows how far absolute retrieval can move when the alignment target changes. CORTIVA uses complementary visual teachers (Wang and Chen, 2026), HCF uses hierarchical intermediate representations (Tang et al., 2026), SAMGA uses subject-aware multi-granularity alignment (Jiang et al., 2026), and NEAR anchors few-repetition queries to stable high-repeat neural structure (Cui et al., 2026). Our question is orthogonal: how much residual alignment capacity is required when the parent target and training pipeline are fixed?

8 Discussion and Limitations

This study establishes alignment capacity as a controllable design variable in EEG-to-vision decoding. A rank-224 residual update removes **41.97%** of the trainable EEG-model parameters while clearing the prospective retention criterion across five independent participants. The compact model also preserves the structure that makes the code useful: ENIGMA–compact geometry remains closely matched, effective dimensionality contracts, correct-image alignment increases slightly, and high-repeat conditioning still steers a fixed image generator toward stimulus-related content. The result therefore establishes capacity allocation as a distinct empirical axis of the decoder.

8.1 Compression reveals a compact visual code

Theorem 1 restricts only the learned residual correction; the direct 1,024-dimensional pathway remains available. Under that constraint, effective rank falls by roughly 9.9%, 10.0%, and 5.4% at $k = 1, 4, 80$, while ENIGMA–compact CKA rises from 0.932 to 0.997 as repetitions increase. Local image-neighborhood overlap is nearly unchanged and correct-image cosine alignment increases at every tested count. This combination is consistent with redundant degrees of freedom in the dense residual correction: the factorization concentrates the representation while preserving the geometry used by retrieval. At $k = 80$, the same compact code remains useful to a separate generation pipeline, extending the result beyond the primary retrieval metric.

The confirmation design strengthens this interpretation. Rank 224, the training rule, and the statistical criterion were frozen before S03–S07 were fitted, so each participant supplies a separate participant-specific replication of the same intervention. The present design establishes prospective participant-specific replication of the capacity effect across five independently optimized encoders; cross-subject transfer defines the next generalization axis.

8.2 Parameter efficiency exposes a second optimization axis

The intervention sharply reduces learned weights while changing executed arithmetic much less: dominant linear/convolutional multiply–accumulate operations (MACs) fall by 4.6%, and recorded A100 training time and batch throughput remain stable. We chose low-rank factorization because it constrains the dimensionality of the learned residual correction while leaving the full identity pathway intact, making residual capacity an identifiable intervention. Structured pruning, low-rank-plus-diagonal maps, LoRA-style residual adapters, spectral regularization, width reduction, and quantization target different structure/compute trade-offs and are directions for reducing execution cost.

8.3 The controlled design creates a clear expansion path

The controlled inference is anchored to an ENIGMA-derived 1,024-dimensional alignment model on THINGS-EEG2. That scope is what makes the capacity intervention identifiable. Rank 224 remains a fixed confirmation point: a subsequent development-only sweep over $r \in \{64, 128, 224, 512\}$ plus dense ENIGMA showed no monotonic retrieval gain with increasing residual rank and placed rank 224 numerically highest in mean development Top-1 (Appendix B). The sweep contextualizes the confirmed operating point without changing the prospective interpretation of the S03–S07 test. Five participants provide participant-level replication, and the fixed 200-image gallery isolates the architectural intervention from stimulus-set changes. Replication on another visual-EEG dataset and target representation would test transport beyond this setting.

The generation experiment supplies a discriminating downstream boundary. At 80 repetitions, both compact-versus-ENIGMA lower bounds clear the DINO-content retention margin; at one trial, the crossed lower bound lies just beyond it even though both models beat shuffled-EEG conditioning. This sharpens the positive high-repeat result and identifies repeat-efficient conditioning as a natural next target. NICE and recent 2026 systems provide complementary context: they show gains from richer visual supervision, hierarchy, subject awareness, and repeat handling, while our matched ENIGMA experiment isolates a different axis. Within that controlled pipeline, almost 42% of trainable EEG-model parameters are removable while confirmed retrieval, representation structure, and high-repeat generative conditioning are retained. Future EEG-to-visual systems can therefore optimize both *what* they align to and *how much learned capacity* they need to carry that alignment.

9 Conclusion

EEG-to-visual decoding is often framed as a problem of learning richer alignments; our results show that **alignment capacity itself deserves the same scrutiny**. By constraining only ENIGMA's final learned residual correction to rank 224, we reduce the trainable EEG-model parameters from **1.405M to 0.815M (41.97%)** while satisfying the prospective single-trial accuracy-retention criterion across five development-independent participants. The compact code remains closely matched to the reference representation, uses fewer effective dimensions, preserves local image-space organization, and slightly increases mean alignment with the correct image target. At 80 repetitions, it also retains stimulus-related conditioning under the fixed image-generation pipeline.

Together, these results identify a compact visual code inside an architecture that originally allocated substantially more residual freedom. The broader implication is methodological: neural decoders should treat learned alignment capacity as a controlled design variable alongside visual targets, data scale, subject adaptation, and repeat handling. This creates two complementary questions for future EEG-to-visual systems: **what visual representation should EEG align to, and how much learned capacity is required to carry that alignment?** In the ENIGMA/THINGS-EEG2 setting studied here, the second answer is substantially smaller than the original dense head allocates. That finding provides a concrete foundation for building neural decoders that are smaller by design while retaining the representation structure and downstream utility that matter for visual decoding.

Reproducibility Statement

The experimental design is specified in Section 3, including the rank-224 intervention, exact parameter counts, training exposure, participant and seed structure, tie-aware retrieval metric, and prospective retention rule. The appendices provide optimizer details, complete repeat-count results, representation-metric definitions, generation settings and controls, NICE reproduction details, and the design-strength/extension matrix. All quantitative claims in the manuscript were traced to saved participant- or image-level outputs during the evidence audit. THINGS-EEG2 is a public dataset. The included supplementary text records the implementation-level settings used for the ENIGMA/compact comparison and the exact public NICE repository commit used for the independent reproduction, together with complete numerical summaries, representation and generation diagnostics, and fixed qualitative grids. Large public inputs and external pretrained weights are not bundled in this manuscript source package.

AI Use Statement

Generative AI tools were used to assist with experimental-protocol refinement and result interpretation; draft and review research code; create and edit scientific figures; verify the completeness and accuracy of bibliographic information against primary sources; and audit the manuscript's quantitative statements, methodological descriptions, and scientific claims against the underlying experimental evidence. AI-assisted code was tested and validated before use, bibliographic information was checked against the cited sources, and reported numerical results were traced to saved experimental outputs. The authors reviewed all AI-assisted work and take responsibility for the final text, claims, code, analyses, figures, and other artifacts.

References

Gifford, A. T., Dwivedi, K., Roig, G., and Cichy, R. M. A large and rich EEG dataset for modeling human visual object recognition. *NeuroImage*, 264:119754, 2022.

Song, Y., Liu, B., Li, X., Shi, N., Wang, Y., and Gao, X. Decoding natural images from EEG for object recognition. In *International Conference on Learning Representations*, 2024.

Bai, Y., Wang, X., Cao, Y.-P., Ge, Y., Yuan, C., and Shan, Y. DreamDiffusion: High-quality EEG-to-image generation with temporal masked signal modeling and CLIP alignment. In *Computer Vision – ECCV 2024*, pp. 472–488. Springer, 2024.

Radford, A. et al. Learning transferable visual models from natural language supervision. arXiv:2103.00020, 2021.

Kneeland, R., Jiang, W., Nunes, U. B., Scotti, P. S., Delorme, A., and Xu, J. ENIGMA: EEG-to-image in 15 minutes using less than 1% of the parameters. arXiv:2602.10361, 2026.

Wang, J. and Chen, K. CORTIVA: Candidate-score fusion of complementary visual teachers for EEG- and MEG-to-image retrieval. arXiv:2608.01355, 2026.

Tang, J., Jiang, S., Su, F., and Zhao, Z. Aligning what EEG can see: Structural representations for brain-vision matching. arXiv:2603.07077, 2026.

Jiang, L., She, Q., Xu, J., Xu, H., Wu, D., and Kuang, Z. Subject-aware multi-granularity alignment for zero-shot EEG-to-image retrieval. arXiv:2604.17782, 2026.

Cui, Z., Kan, S., Liu, D., and Wu, D. Beyond trial averaging: Anchoring neural and visual representations for few-repetition brain-to-image retrieval. arXiv:2608.19128, 2026.

Banville, H., Benchetrit, Y., d'Ascoli, S., Rapin, J., and King, J.-R. Scaling laws for decoding images from brain activity. arXiv:2501.15322, 2025.

Denton, E., Zaremba, W., Bruna, J., LeCun, Y., and Fergus, R. Exploiting linear structure within convolutional networks for efficient evaluation. arXiv:1404.0736, 2014.

Kornblith, S., Norouzi, M., Lee, H., and Hinton, G. E. Similarity of neural network representations revisited. arXiv:1905.00414, 2019.

A Supplementary Methodological Details

This appendix records the implementation details that support Section 3 while keeping the main paper focused on the capacity intervention and prospective confirmation.

A.1 Data contract

The prepared THINGS-EEG2 training arrays contain 16,540 image identities, four EEG recordings per image, 63 channels, and 250 samples per recording. The corresponding test arrays contain 200 image identities with 80 recordings per image. Training and test image identities are disjoint in the prepared benchmark. The model consumes the fixed preprocessed arrays used throughout the ENIGMA replication pipeline; image and text target features remain fixed during EEG training.

The source-matched exposure contract uses the first 66,048 of the 66,160 available flat training trials in every epoch. These trials form 129 complete batches of 512. Both the reference and compact models receive the same per-epoch shuffled trial plan for a matched participant and seed.

A.2 Training objective and optimizer

For an EEG output z_i, fixed image target v_i, and fixed text target t_i, training minimizes

$$\mathcal{L} = \mathrm{MSE}(z_i, v_i) + 0.5\,\mathcal{L}_{\mathrm{img}} + 0.05\,\mathcal{L}_{\mathrm{text}}.$$

The two contrastive terms are symmetric in-batch cross-entropies. Target vectors are unit-normalized inside the contrastive logits; EEG outputs remain at their learned scale, and the trainable logit scale is used directly. AdamW uses a maximum learning rate of 3×10^{-4}, source-matched weight-decay groups of 0.01 and 0, and a OneCycleLR schedule with `pct_start=0.1` and `final_div_factor=1000`. Gradients are clipped to global norm 1. Each model receives 50 epochs and 6,450 optimizer steps. The fixed epoch-50 model supplies all confirmation results. Core training uses float32 operations with TF32 enabled and deterministic cuDNN settings.

A.3 Parameter accounting

The ENIGMA residual matrix contains

$$1024^2 = 1{,}048{,}576$$

weights. The rank-224 factorization contains

$$1024(224) + 224(1024) = 458{,}752$$

weights, a reduction of 589,824 or 56.25% within the residual matrix. The complete reference and compact EEG models contain 1,405,243 and 815,419 trainable parameters, respectively. Their difference is exactly the 589,824 weights removed from the residual transformation, corresponding to a 41.9731% reduction at model level.

A.4 Tie-aware retrieval metrics

For a query whose correct candidate has g candidates with strictly larger score and shares its score with a tie block of size e, the possible ranks of the correct candidate are

$$g + 1, \ldots, g + e.$$

Metric credit is averaged uniformly over those ranks. For example, Top-1 credit is $1/e$ when $g = 0$ and zero when $g > 0$. If all 200 candidates are exactly tied, Top-1 therefore equals $1/200 = 0.005$. Every primary score row is also checked by an independently implemented sorted-boundary evaluator.

A.5 Repeat aggregation

The primary endpoint uses one EEG recording at a time. Secondary analyses use $k \in \{2, 4, 8, 16, 32, 80\}$ real repetitions under two fixed aggregation paths. Waveform-first aggregation averages the selected EEG recordings before encoding. Representation-first aggregation encodes each selected recording and averages the resulting representations. For $k < 80$, 50 deterministic nested subsets per image are used; $k = 80$ uses all available repetitions once.

A.6 Prospective retention analysis

For the primary single-trial endpoint, trial-level outcomes are averaged within each of the 200 image clusters. The two training seeds are then averaged within participant. Let Δ_s denote the resulting compact-minus-reference effect for participant s. The prespecified retention margin is $m = 0.005$ absolute accuracy, or 0.5 percentage points. This threshold was inherited from the development plan as a fixed practical decision rule for authorizing a capacity reduction; it was not estimated from S03–S07.

Rank 224 likewise originated before confirmation in the B10 capacity-allocation study. There, on development participants S01–S02 with seeds 42 and 43, dense ENIGMA achieved 6.803% mean single-trial Top-1, while the rank-224 compact-only arm achieved 7.067% with 41.97% fewer model parameters; widening-only and widened-plus-compact arms reached 6.048% and 6.125%, respectively. We therefore froze the already specified rank-224 compact head for confirmation rather than tuning r on S03–S07.

The participant analysis reports a one-sided 95% lower t bound across the five participant effects. A complementary crossed bootstrap uses 10,000 draws in which participants and the shared 200 image clusters are resampled; the same sampled image weights are applied to both models and across participants. Retention requires both one-sided lower bounds to exceed $-m$ and reference-model point accuracy to exceed chance in every fit. Superiority is evaluated only after retention, requiring both lower bounds above zero and a one-sided exact sign-flip $p \leq 0.05$ over the five participant effects.

This design treats participants as the population-level replication units while preserving the paired image structure within the fixed 200-image gallery.

The confirmed one-sided lower bounds are -0.106 pp (participant-t) and -0.131 pp (crossed participant/image). Consequently, the retention decision is unchanged throughout the reviewer-suggested loss-margin range of 0.25–1.0 pp: both bounds clear margins of -0.25, -0.50, -0.75, and -1.00 pp. Using the unrounded crossed bound, both analyses would still pass for any loss margin of approximately 0.132 pp or larger.

B Development-Only Residual-Rank Sweep

B.1 Origin of the rank-224 operating point

Rank 224 was not selected from an accuracy-based rank sweep. It originated as a parameter-budget operating point in the development-stage capacity-allocation design. In that design, factorizing the 1024×1024 residual matrix as 1024×224 and 224×1024 allowed a widened 16-map EEG encoder to remain at 1,381,159 trainable parameters, slightly below the 1,405,243-parameter dense ENIGMA reference. The same rank was also evaluated in the compact-only arm. On development participants S01–S02 with seeds 42 and 43, that compact-only arm achieved 7.067% mean 200-way single-trial Top-1 versus 6.803% for dense ENIGMA. The rank-224 compact architecture was therefore frozen before the S03–S07 confirmation study; no accuracy-based rank sweep or confirmation outcome was used to select r.

B.2 Subsequent development-only rank sweep

To contextualize the fixed operating point, we subsequently evaluated residual ranks $r \in \{64, 128, 224, 512\}$ together with dense ENIGMA on the same development participants S01–S02 and seeds 42 and 43. The dense and rank-224 results reuse the original development fits; ranks 64, 128, and 512 were trained under the same 50-epoch training and 200-way single-trial retrieval protocol. This analysis is development-only and does not alter the prospective S03–S07 confirmation claim.

The sweep does not show a monotonic benefit from increasing residual rank. Rank 224 has the highest mean development Top-1 at 7.067%, but rank 64 is only 0.014 percentage points lower at 7.053% despite using 65.29% fewer parameters than dense ENIGMA. The tested range therefore supports substantial residual-capacity redundancy but does not identify a minimum sufficient rank or a sharp failure boundary.

Table 3: **Development-only residual-rank sweep.** Top-1 values are means across the four participant–seed cells from S01–S02 and seeds 42 and 43. Cell SD is reported descriptively in percentage points. Parameter reduction is relative to dense ENIGMA.

Model	Residual rank	EEG params	Reduction (%)	Top-1 (%)	Cell SD (pp)
Rank-constrained	64	487,739	65.29	7.053	0.191
Rank-constrained	128	618,811	55.96	6.892	0.332
Rank-constrained	224	815,419	41.97	**7.067**	0.141
Rank-constrained	512	1,405,243	0.00	6.737	0.156
Dense ENIGMA	full	1,405,243	0.00	6.803	0.108

Participant-specific results likewise do not identify one universal optimum. After averaging the two seeds, S01 is numerically highest at $r = 128$ (7.053%), whereas S02 is highest at $r = 64$ (7.200%). Rank 224 remains near the top for both participants, at 7.034% for S01 and 7.100% for S02. These development-only results motivate future participant-adaptive capacity selection, while the prospective confirmation claim in the main paper remains restricted to the frozen rank-224 model evaluated on S03–S07.

C Supplementary Retrieval Results

This appendix expands the controlled confirmation results summarized in Section 4. Seeds are averaged within participant before participant-level summaries.

C.1 Participant-level single-trial effects

Table 4: Seed-averaged 200-way single-trial Top-1 by confirmation participant. Differences are compact minus ENIGMA in percentage points.

Participant	ENIGMA (%)	Compact (%)	Difference (pp)
S03	9.597	9.512	-0.084
S04	9.012	9.000	-0.012
S05	5.966	6.041	+0.075
S06	9.266	9.378	+0.112
S07	10.075	10.612	+0.537

The participant effects are -0.084, -0.012, +0.075, +0.112, +0.537 percentage points for S03–S07. The one-sided 95% participant-t lower bound is -0.106 pp. The crossed participant/image bootstrap lower bound is -0.131 pp, based on 10,000 draws over five participants and the shared 200-image gallery. The corresponding two-sided 95% intervals are [-0.176, +0.427] pp and [-0.182, +0.446] pp. The one-sided exact participant sign-flip test gives $p = 0.1875$.

C.2 Complete repeat-count Top-1 summaries

Table 5: Representation-first averaging across the prespecified number of real EEG repetitions.

k	ENIGMA (%)	Compact (%)	Difference (pp)
1	8.783	8.909	+0.126
2	13.580	13.576	-0.004
4	19.040	19.066	+0.026
8	23.773	23.831	+0.058
16	27.133	27.058	-0.075
32	29.003	29.094	+0.091
80	29.650	30.200	+0.550

Table 6: Waveform-first averaging across the prespecified number of real EEG repetitions.

k	ENIGMA (%)	Compact (%)	Difference (pp)
1	8.783	8.909	+0.126
2	13.261	13.370	+0.109
4	17.976	18.063	+0.087
8	21.769	21.980	+0.211
16	24.679	24.758	+0.079
32	26.232	26.267	+0.035
80	27.350	27.200	-0.150

The prospective retention decision is attached to the single-trial endpoint. The repeat-count tables report the complete fixed secondary curves.

C.3 Measured training and inference cost

Table 7: Measured cost summaries across the ten fits per architecture. Parameter reduction yields a modest change in dominant linear/convolutional MACs, while recorded A100 training time and throughput remain essentially unchanged.

Model	MACs/trial	Train (min)	Batch-1 (ms)	Batch-128 trials/s
ENIGMA	12,918,252	8.70	0.749	53988
Compact	12,328,428	8.69	0.827	53784

Mean peak CUDA allocation is 2.812 GiB for ENIGMA and 2.803 GiB for compact. These measurements support the manuscript's parameter-efficiency claim while keeping runtime behavior as a separately measured quantity.

D Representation Analysis Details

The representation analysis uses the final raw 1,024-dimensional outputs saved for both models. Counts $k \in \{1, 4, 80\}$ share a deterministic repeat plan. Unit normalization is applied only for cosine-based analyses and effective-rank calculation. Both training seeds enter the original-space statistics; encoder seed 42 supplies the fixed UMAP displays.

D.1 Metric definitions

For centered representation matrices X and Y with matching image order, linear CKA uses the sample Gram matrices $K = XX^\top$ and $L = YY^\top$:

$$\mathrm{CKA}(X, Y) = \frac{\langle K, L \rangle_F}{\|K\|_F \|L\|_F}.$$

A 20-permutation shuffled-image reference is computed for the EEG–image CKA values.

Correct-image cosine alignment is the mean dot product between unit-normalized EEG representation $\hat{z}_i$ and its corresponding unit-normalized image representation $\hat{v}_i$:

$$A_{\cos} = \frac{1}{N} \sum_{i=1}^{N} \hat{z}_i^\top \hat{v}_i.$$

For neighborhood preservation, each EEG representation and image representation defines a ten-nearest-neighbor set over the same 200 image identities, with the query identity excluded. The per-image score is

$$O_i = \frac{|\mathcal{N}_{10}^{\mathrm{EEG}}(i) \cap \mathcal{N}_{10}^{\mathrm{img}}(i)|}{10}.$$

Effective rank is computed from the singular values s_j of the centered, unit-normalized representation matrix:

$$p_j = \frac{s_j^2}{\sum_\ell s_\ell^2}, \qquad r_{\text{eff}} = \exp\left(-\sum_j p_j \log p_j\right).$$

D.2 Complete original-space summaries

Table 8: Representation metrics averaged across five confirmation participants and two encoder seeds.

k	CKA B–C	CKA B–img	CKA C–img	Cos B	Cos C	10-NN B	10-NN C	Rank B	Rank C
1	0.9320	0.3257	0.3178	0.0987	0.1047	0.0872	0.0873	61.83	55.69
4	0.9742	0.3172	0.3130	0.1456	0.1518	0.1238	0.1249	39.39	35.43
80	0.9967	0.2984	0.2987	0.1892	0.1944	0.1565	0.1582	16.57	15.67

Here B denotes the ENIGMA reference and C the compact model.

D.3 Paired uncertainty for alignment and neighborhoods

Seeds are averaged before participant-level uncertainty calculations. Participant-t intervals use the five participant effects. Crossed intervals resample participants and the shared image identities jointly.

Table 9: Compact-minus-ENIGMA paired differences for correct-image cosine alignment and ten-neighbor overlap. These are secondary descriptive intervals.

k	Cosine diff.	Cosine t 95% CI	Cos. crossed 95% CI	10-NN diff.	10-NN t 95% CI	10-NN crossed 95% CI
1	+0.00594	[+0.00438, +0.00750]	[+0.00427, +0.00761]	+0.00010	[-0.00447, +0.00467]	[-0.00500, +0.00515]
4	+0.00620	[+0.00439, +0.00802]	[+0.00459, +0.00774]	+0.00115	[-0.00448, +0.00678]	[-0.00445, +0.00630]
80	+0.00514	[+0.00247, +0.00782]	[+0.00324, +0.00678]	+0.00175	[-0.00128, +0.00478]	[-0.00160, +0.00500]

D.4 UMAP display protocol

The UMAP input contains 4,200 vectors: ENIGMA and compact outputs for five participants at $k = 1$ and $k = 80$ using encoder seed 42, together with the 200 matching image vectors. The projection uses cosine distance, 15 neighbors, minimum distance 0.15, 400 epochs, two output dimensions, and one thread. Object labels do not enter fitting. Two fixed map seeds are retained: seed 42 has fixed-sample trustworthiness 0.876, and seed 43 has trustworthiness 0.873. Figure 3 uses map seed 42; the alternate seed and the target-inclusive maps are retained below.

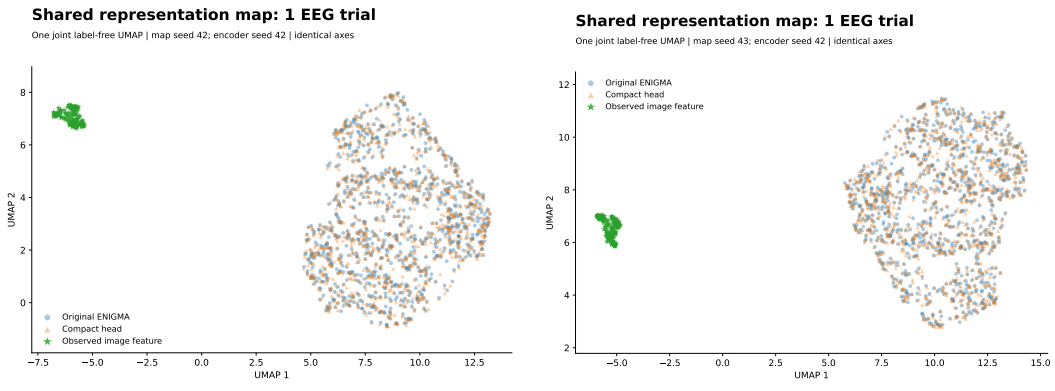

Figure 5: Shared UMAP at one EEG repetition under the two fixed map seeds. The same high-dimensional input vectors are projected independently with random states 42 and 43.

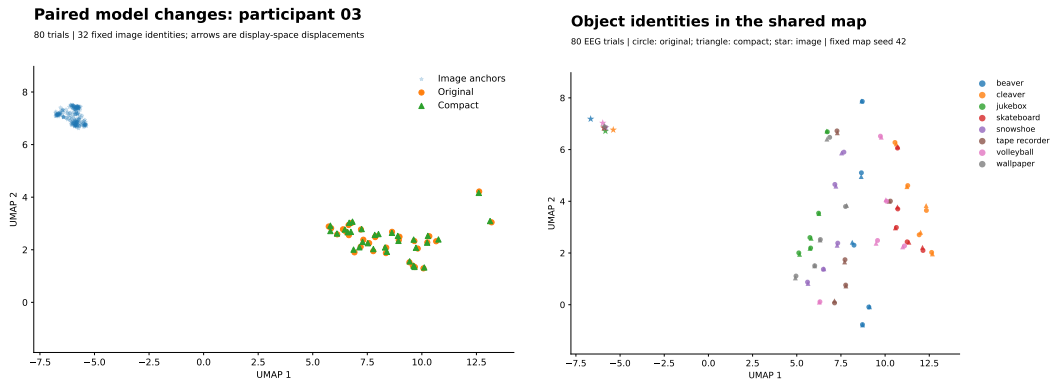

Figure 6: Additional shared-coordinate views at 80 repetitions. Left: fixed paired ENIGMA-to-compact endpoint displacements for participant S03. Right: eight fixed object identities shown with model and image markers in the same map. These panels are display-space summaries; original-space metrics carry the quantitative interpretation.

E Downstream Generation Details

The downstream experiment conditions a fixed SDXL-Turbo generator through the standard pooled ViT-H IP-Adapter using the raw 1,024-dimensional EEG-derived image representation. Generation uses an empty prompt, four diffusion steps, 512-pixel output, guidance scale 0, IP-Adapter scale 1, and one fixed diffusion-noise seed per stimulus. The same noise seed is used across ENIGMA, compact, shuffled, oracle, and zero-conditioning variants for that stimulus.

E.1 Fixed stimulus and control design

Thirty-two of the 200 test identities are selected once with seed 20260923. All five confirmation participants contribute ENIGMA, compact, ENIGMA-shuffled, and compact-shuffled conditions at $k = 1$ and $k = 80$. Encoder seed 42 is fixed for this stage. Shuffling maps each target identity to another image identity before conditioning. Oracle conditioning uses the known image embedding; zero conditioning uses a zero vector. Across all conditions the stage contains 1,344 generated images.

The oracle-minus-zero DINO cosine control has a one-sided bootstrap lower bound of 0.4995, confirming that the fixed generator responds strongly to the available conditioning channel.

E.2 Content metric and retention rule

Generated image $\hat{I}_i$ is compared with its observed stimulus I_i using frozen DINOv2-small CLS features:

$$S_{\mathrm{DINO}}(\hat{I}_i, I_i) = \hat{f}(\hat{I}_i)^\top \hat{f}(I_i).$$

For each EEG model and repetition count, correctly conditioned similarity must exceed shuffled-EEG similarity under both the participant-based and crossed participant/image one-sided lower bounds before model-to-model retention is interpreted.

The compact-minus-ENIGMA practical retention margin is 0.02 cosine units. Retention requires both one-sided 95% lower bounds to exceed -0.02.

The compact-minus-ENIGMA cosine difference is -0.0069 at $k = 1$ and +0.0050 at $k = 80$. Corresponding participant-t lower bounds are -0.0150 and -0.0165; crossed participant/image lower bounds are -0.0205 and -0.0153. The $k = 80$ condition satisfies the predefined retention rule.

Table 10: Generation-content summary. DINO Top-1/Top-5 are 200-way retrieval scores computed from generated-image features against the fixed stimulus gallery.

Condition	DINO cosine	Top-1 (%)	Top-5 (%)	Corr.–shuffled	t lower	Crossed lower
$k = 1$ ENIGMA	0.0445	1.875	8.750	0.032	+0.0239	+0.0158
$k = 1$ Compact	0.0376	1.875	6.250	0.026	+0.0150	+0.0074
$k = 80$ ENIGMA	0.0769	6.250	20.000	0.055	+0.0437	+0.0326
$k = 80$ Compact	0.0819	8.125	23.750	0.059	+0.0397	+0.0334

E.3 Complete fixed example grids

The following pages reproduce the complete prespecified participant-S03 image grids. Each page contains eight fixed stimulus identities; four pages cover all 32 identities. The ordering comes from the seed-fixed selection and is retained without visual filtering.

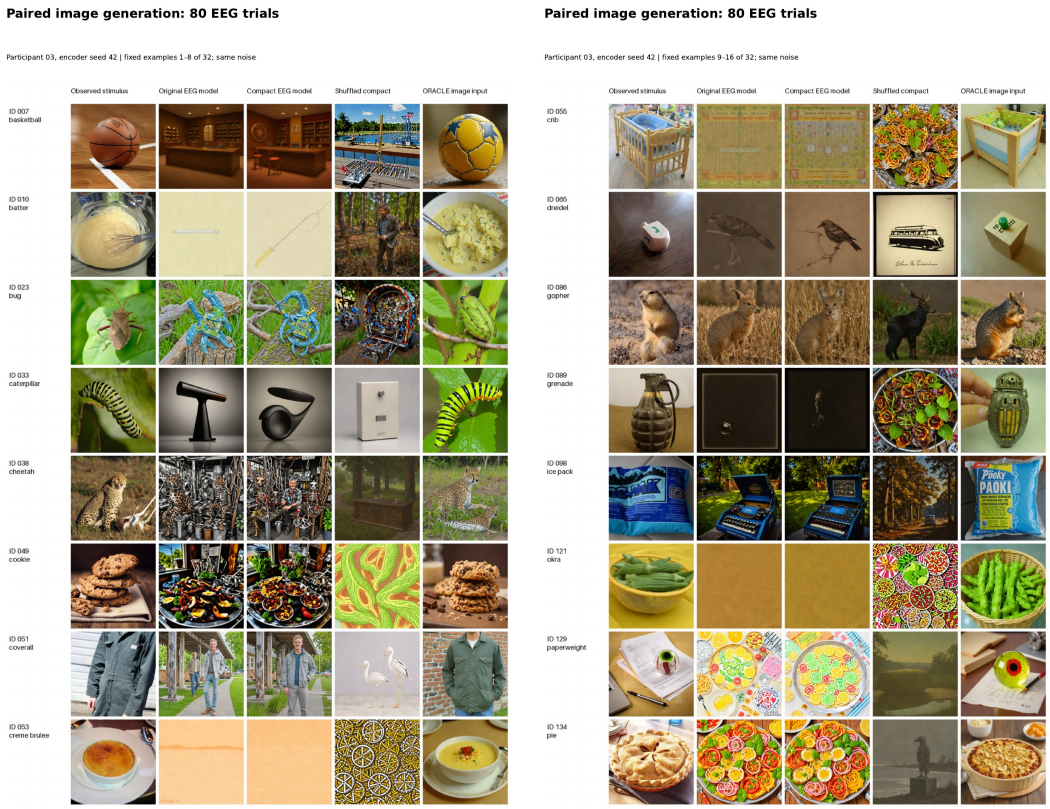

Figure 7: High-repeat generation examples, fixed identities 1–16 of 32.

F NICE Reproduction and External-Context Details

The final baseline experiment reproduces NICE from the official public repository at commit `a7709d81af9c4971c711b03fbee4e09e01477c4b`. The five confirmation participants S03–S07 are trained separately with seeds 42 and 43, yielding ten NICE fits. The public NICE CLIP-768 visual targets and public model construction are retained.

F.1 Reproduction protocol

For each participant, the four training repetitions are averaged for each of 16,540 training images and the 80 test repetitions are averaged for each of 200 test images. A seed-specific permutation



Figure 8: High-repeat generation examples, fixed identities 17–32 of 32.

reserves 740 training images for validation and uses the remaining 15,800 for fitting. NICE is trained for 200 epochs with batch size 1000, Adam learning rate 2×10^{-4}, betas $(0.5, 0.999)$, and symmetric contrastive cross-entropy on normalized EEG and CLIP-768 features. The checkpoint with minimum validation contrastive loss is selected before test evaluation.

The public construction contains image-projector weights whose forward path returns the image features unchanged. These unused projector weights remain in the optimizer to preserve the public construction and random-number path, while the comparable EEG inference parameter count is 1,802,496.

F.2 Participant-level NICE results

Table 11: Seed-averaged NICE Top-1 and compact ENIGMA $k = 80$ waveform Top-1 on the same participant/test-image identities. The final column is descriptive because the visual targets and training protocols differ.

Participant	NICE (%)	Compact (%)	Difference (pp)
S03	12.50	30.25	+17.75
S04	17.50	32.00	+14.50
S05	6.50	19.25	+12.75
S06	14.25	27.25	+13.00
S07	15.50	27.25	+11.75

Across participants, the compact-minus-NICE Top-1 difference averages +13.95 percentage points. The participant 95% interval is [+11.04, +16.86] pp, and the crossed participant/image interval is [+9.05, +18.80] pp.

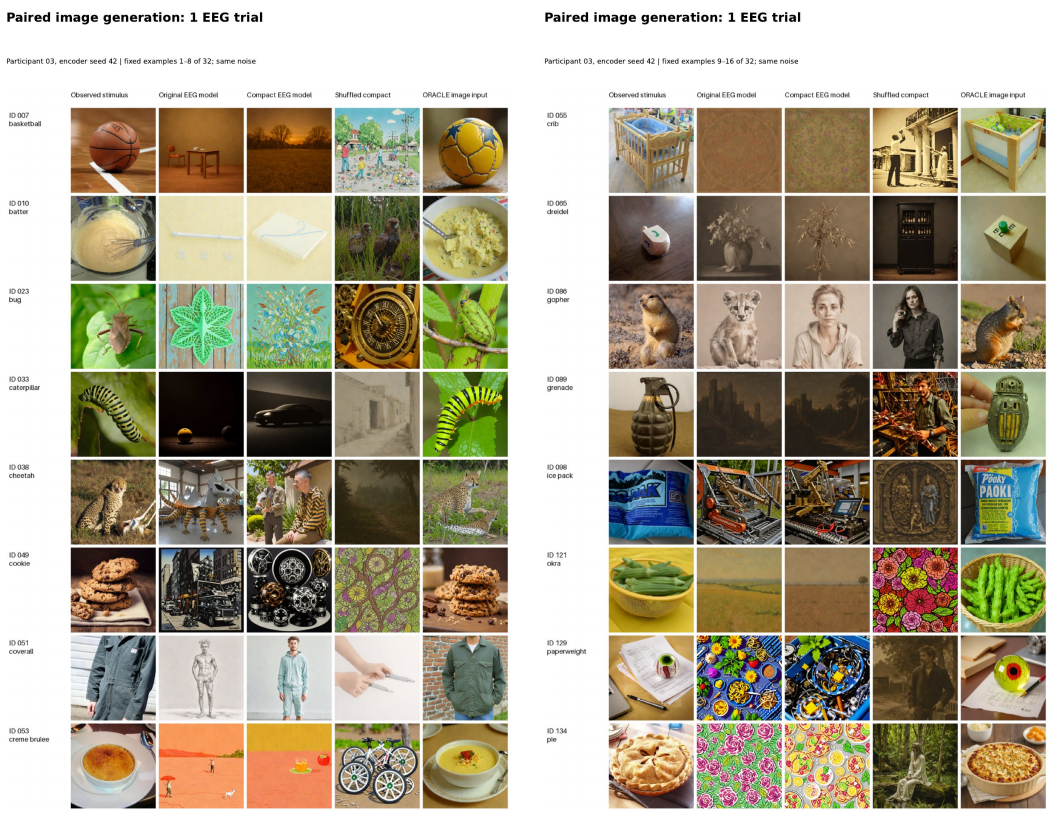

Figure 9: Single-trial generation examples, fixed identities 1–16 of 32.

F.3 Protocol boundary for external comparisons

The exact parent-model comparison in the main paper is ENIGMA reference versus compact ENIGMA under the same 1,024-dimensional target, training samples, optimizer schedule, seeds, and evaluation procedure. NICE provides an independent architecture benchmark with the same THINGS-EEG2 participant/test-image identities while using its public CLIP-768 target and mean-repetition training. Published CORTIVA, HCF, SAMGA, and NEAR values are transcribed from their papers and retain each method's visual targets, alignment strategy, and evaluation design.

G Design Strengths and Expansion Tests

The main experiments were constructed to isolate the capacity intervention while preserving a direct route for broader confirmation. Table 12 summarizes how each design choice strengthens the current inference and which experiment most directly extends it.

H Extended Related-Work Comparison

Table 13 separates the main optimization axes represented in the EEG-to-visual literature. Its purpose is methodological: published systems vary visual supervision, subject handling, data scale, and repetition strategy simultaneously, whereas the present study isolates downstream residual capacity inside one fixed parent pipeline.



Figure 10: Single-trial generation examples, fixed identities 17–32 of 32.

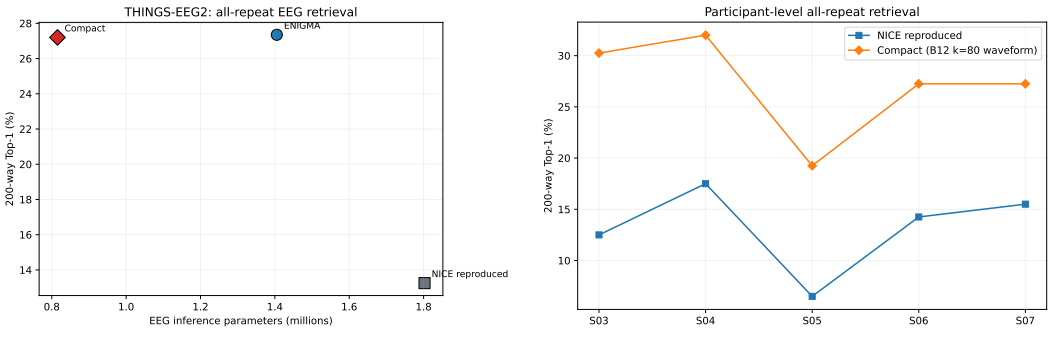

Figure 11: NICE reproduction diagnostics. Left: measured accuracy and EEG inference parameter counts for reproduced NICE and the ENIGMA-derived models under their respective protocols. Right: participant-level NICE and compact results across S03–S07.

Table 12: Design choices that strengthen the present inference and the highest-value extension enabled by each one.

Design choice	Strength added to the present study	Highest-value extension
Rank 224 frozen before confirmation	Turns the compact architecture into a prospective test rather than a post hoc compression choice	Extend the development sweep below rank 64 and across additional development participants, then prospectively confirm any newly selected operating point on a separate cohort
Five development-independent participants with separate fits	Makes each participant a genuine replication unit for the capacity intervention	Scale the same frozen architecture to a larger cohort and a second visual-EEG dataset
Fixed 200-image gallery and matched ENIGMA training pipeline	Holds stimulus composition, visual target, optimization, and retrieval protocol constant so the architectural intervention remains isolated	Repeat the same capacity test under a second target representation while preserving matched training exposure
Representation analysis in the original 1,024-D space	Links retained retrieval to preserved geometry, alignment, neighborhoods, and effective dimensionality	Relate residual directions to visual attributes, EEG time windows, or electrode subsets through targeted interventions
Prespecified generation controls and retention margin	Shows that the downstream criterion can distinguish regimes: $k = 80$ passes while the noisier single-trial regime sits at the boundary	Improve repeat-efficient conditioning and validate with multiple diffusion seeds, generators, and human perceptual scoring
Independent NICE reproduction plus current published methods	Anchors the compact model outside the parent architecture while preserving protocol provenance	Build a fully protocol-matched multi-architecture study with one shared visual target and one shared training/evaluation contract
Measured parameters, MACs, timing, throughput, and peak allocation	Separates learned-capacity reduction from execution-path reduction and identifies two independent optimization axes	Combine the validated rank-constrained head with width reduction, structured pruning, quantization, or deployment-specific optimization

Table 13: Representative EEG-to-visual systems organized by the modeling axis they change. Published methods retain their own targets and evaluation protocols.

Work	Task	Representation strategy	Primary axis varied	Relation to this study
THINGS-EEG2 (Gifford et al., 2022)	Benchmark	Repeated natural-image EEG	Data scale and standardized visual-EEG evaluation	Public confirmation setting
NICE (Song et al., 2024)	Retrieval	Contrastive EEG–image alignment	Shared visual embedding and zero-shot retrieval	Establishes the retrieval formulation
DreamDiffusion (Bai et al., 2024)	Generation	Temporal EEG modeling, CLIP alignment, diffusion	Generative conditioning	Shows that aligned EEG codes can condition pretrained generators
Scaling laws (Banville et al., 2025)	Benchmark	EEG/MEG/fMRI image decoding	Neural data amount and recording modality	Characterizes the data-scaling axis
ENIGMA (Kneeland et al., 2026)	Retrieval / generation	Compact multi-subject EEG-to-visual architecture	Subject adaptation and model-wide parameter efficiency	Parent architecture for the controlled intervention
CORTIVA (Wang and Chen, 2026)	Retrieval	Complementary visual teachers	Candidate-score fusion	Strengthens visual supervision
HCF (Tang et al., 2026)	Retrieval	Intermediate visual representations	Hierarchical complementary fusion	Strengthens structural alignment
SAMGA (Jiang et al., 2026)	Retrieval	Subject-aware multi-granularity targets	Subject and target granularity	Strengthens subject-conditioned alignment
NEAR (Cui et al., 2026)	Retrieval	High-repeat neural anchors	Few-repetition inference	Strengthens repeat-efficient retrieval
This study	Capacity test	Fixed ENIGMA target, training, and evaluation	Rank of the learned residual correction	Isolates downstream alignment capacity