IDENTITY RECOVERY FROM REPEATED CAMPRO OUTPUTS UNDER SYNTHETIC ACQUISITION NOISE

Anonymous authors

ABSTRACT

Privacy-oriented imaging aims to suppress facial recognition, but repeated releases may support identity recovery. We examine identity recovery from repeated outputs of the CamPro image signal processor (ISP), using independent synthetic pre-ISP noise and fixed, aligned face images. Comparisons include separately selected single-output attacks, reconstruction averaging, joint inversion and matched pixel-mean inversion, alongside color-channel and fixed-weight detection analyses. Eight observations improve FaceScrub closed-gallery Rank-1 by 14.6–20.3 percentage points over the selected single-output attack; NTU gains are less consistent. The FaceScrub gain persists after 8-bit lossless export, while strict open-set acceptance remains limited under the tested calibration. A 500-identity synthetic study reports mean gains of 5.9–9.2 points across recognizers over three independently trained and selected fits; averaging recovers much of this gain. Joint inversion outperforms matched pixel-mean inversion on FaceScrub. Eight of twelve constructed color pairs with identical noiseless outputs have distinguishable noisy output distributions. Added noise lowers native person AP from 0.518 to 0.505; task and identity inputs use different spatial preprocessing. These results motivate repeated-release privacy evaluation and show that deterministic ambiguity need not survive random acquisition.

1 INTRODUCTION

Camera-based privacy protection aims to limit identity exposure while retaining images for non-sensitive tasks. CamPro implements this approach in the image signal processor (ISP), modifying color correction and tone mapping to suppress facial recognition while supporting person detection (Zhu et al., 2024). Released images constitute the observer's measurements, so privacy evaluation must specify how many outputs the observer can access.

Multiple observations can improve image restoration, as established in burst restoration (Delbracio & Sapiro, 2015; Aittala & Durand, 2018). For the specified privacy-oriented ISP, the question is how much of this improvement carries over to identity recovery. We held the aligned source image fixed, introduced independent synthetic RGB perturbations before the same static ISP, and gave the attacker nested sets of protected images. This protocol examines repeated measurements of the same source without introducing the additional poses or context available in natural video. We also tested whether colors merged by the deterministic ISP remained indistinguishable after pre-ISP perturbation.

We compared three uses of the released measurements: single-output inversion, averaging of single-output reconstructions, and joint set inversion. These comparisons distinguish gains from additional observations from those associated with joint processing. The feature-moment architecture in Figure 1 serves as a variable-cardinality baseline. We also fitted inverses to noiseless outputs and to the pixel mean of eight noisy observations, complementing the late-fusion comparison with input-domain matching. A fixed-weight person-detection evaluation measured the task effect of the same perturbation definition at the detector's native input resolution.

On FaceScrub, joint inversion of eight outputs improved closed-gallery Rank-1 by 14.6–20.3 percentage points over the independently selected single-output attack; NTU gains were less consistent. In the separate DigiFace study, gains persisted across three training seeds, but averaging independent reconstructions recovered much of the improvement. The FaceScrub effect also persisted after 8-bit lossless export. Open-set evaluation showed that improved ranking did not necessarily yield

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.

acceptance at a strict threshold in the evaluated gallery. These comparisons concern the specified attack budgets. The color analysis identified distinguishable noisy output distributions for inputs merged by the noiseless mapping, but did not quantify this effect's contribution to the measured identity gains.

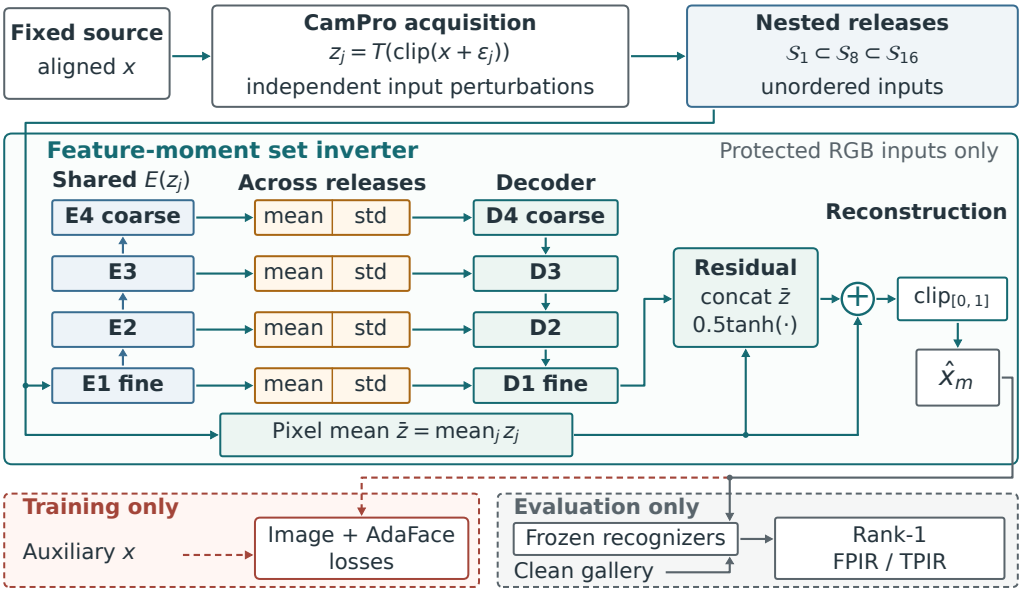


Figure 1: Set inversion of repeated CamPro outputs. Independent perturbations precede the same static ISP. Shared multiscale features are pooled by their mean and stabilized standard deviation. The observation pixel mean both conditions the residual head and bypasses it to the final sum and clip. Clean auxiliary images supervise training only. Table 1 compares one-output inversion, averaging eight independent reconstructions, and this joint path. Selection and threshold calibration use development data.

2 Related Work

Privacy-oriented imaging. Optical filtering can preserve selected task signals before digitization (Pittaluga & Koppal, 2015; Hinojosa et al., 2021; Lopez et al., 2024). DyPP jointly optimizes spectral noninvertibility, diversity, privacy and task utility in a dynamic optical model (Cheng et al., 2024). CamPro instead modifies a static ISP and evaluates person detection together with facial-recognition resistance, including restoration and retrained recognizers (Zhu et al., 2024). We study repeated outputs of its released digital mapping under a declared synthetic perturbation model, not a calibrated hardware acquisition process.

Identity across observations. Face-set recognition combines multiple images through pooling, quality weighting or learned aggregation (Hassner et al., 2016; Yang et al., 2017; Liu et al., 2017; Kim et al., 2022b; Jawade et al., 2023). Identity can also remain recognizable through body and contextual cues after face obfuscation (Oh et al., 2016). Our repeated-measurement protocol adds neither new clean views nor contextual cues. All members originate from the same source image.

Restoration and set models. Fourier accumulation and tied convolutional networks recover information from bursts (Delbracio & Sapiro, 2015; Aittala & Durand, 2018). Deep Sets and Set Transformers provide symmetric aggregation (Zaheer et al., 2017; Lee et al., 2019). BIPNet and Burstormer add alignment and attention (Dudhane et al., 2022; 2023). CodeFormer and DiffBIR introduce learned face and diffusion priors (Zhou et al., 2022; Lin et al., 2024). Our comparisons evaluate biometric recovery, which can differ from perceptual or pixel-level restoration quality. Pretrained postprocessing and mechanism-fitted inversion remain distinct attack conditions.

Noise before nonlinear measurements. Dithering and noise-enhanced estimation study how noise preceding a nonlinear operation can improve estimation under specified conditions (Chen et al., 2008). We examine this possibility in a privacy-oriented ISP through exact noiseless color collisions and their noisy output distributions.

3 Threat Model and Comparisons

Let x be a clean aligned face image and T the fixed CamPro ISP. The capture pipeline produces

$$z_j = T(\mathrm{clip}(x + \epsilon_j, 0, 1)), \qquad \epsilon_j \sim \mathcal{N}(0, \sigma^2 I), \tag{1}$$

with independent draws and $\sigma = 0.01$ in the primary condition. The color transform and clipping operations are specified in Section 5. The attacker knows this mapping and distribution, but not the realized perturbations or the clean test query. It fits an inverse on paired auxiliary identities and computes

$$\hat{x}_m = A(\mathcal{S}_m), \qquad \mathcal{S}_m = \{z_1, \ldots, z_m\}. \tag{2}$$

The unordered multiset retains repeated members. Its prefixes are nested, so increasing m adds observations without replacing earlier ones. Each of an identity's four source queries defines a separate set. The latent image is fixed within a set; natural motion is outside this primary benchmark.

For a frozen recognizer r, closed-gallery Rank-1 requires the correct identity to rank above all other enrolled identities. We report absolute rates and two distinct differences. For the joint model J, held fixed across input counts,

$$\Delta_r(m) = R_r(J, m) - R_r(J, 1). \tag{3}$$

The principal comparison instead uses a separately selected single-output attack S and reports $R_r(J, 8) - R_r(S, 1)$. Averaging eight reconstructions from S supplies the late-fusion comparison. All paired methods receive the same underlying observations. Human selections use S_{e} from the expanded candidate pool; S_{d} denotes the dedicated-budget pool used in supporting comparisons. For CamPro both choose the same model. Synthetic and export-adapted models are $S_{\mathrm{s}}, J_{\mathrm{s}}$ and $S_{\mathrm{x}}, J_{\mathrm{x}}$. Selection uses development identities, separately from evaluation. A_0 is fitted to noiseless outputs, while B_8 is fitted to the pixel mean of eight noisy outputs, computed before the encoder.

An optimal attacker can discard extra observations, so

$$R_r^\star(m') \geq R_r^\star(m) \quad \text{for } m' > m. \tag{4}$$

This property of an optimal attacker guarantees neither monotonicity nor a particular gain for a fixed learned estimator at a finite training budget. At $\sigma = 0$, a static ISP returns identical outputs. Noise before a nonlinear ISP changes the observation distribution. Our experiments measure recovery from repeated draws. A separate color experiment tests whether the random channel preserves the deterministic mapping's input ambiguities.

4 Set Inversion of CamPro Outputs

Each z_j passes independently through one shared four-scale convolutional encoder with channel widths 32, 64, 96 and 128. Its 3×3 convolutional blocks use group normalization, SiLU and residual connections, with stride two between scales. Let $f_{j\ell}$ be its feature tensor at scale ℓ. We aggregate each scale with symmetric moments

$$a_\ell(\mathcal{S}_m) = \left[\frac{1}{m} \sum_{j=1}^{m} f_{j\ell}, \sqrt{\frac{1}{m} \sum_{j=1}^{m} (f_{j\ell} - \bar{f}_\ell)^2 + \varepsilon} \right]. \tag{5}$$

Here $\bar{f}_\ell = m^{-1} \sum_j f_{j\ell}$ and $\varepsilon = 10^{-6}$. At $m = 1$ the standard-deviation channels are constant 10^{-3}. We select the single-output attack separately on development data using one unique observation per query. The symmetric reduction is permutation invariant in exact arithmetic; double precision limits reduction error. Mean and standard deviation summarize shared evidence and perturbation-induced variation. For a fixed encoder and independent clipped captures, they converge to the corresponding mechanism-conditioned moments.

A decoder with a 192-channel bottleneck and three bilinear upsampling stages fuses a_ℓ from coarse to fine resolutions. After the finest stage, the output head concatenates the decoded feature with the pixel mean $\bar{z}$ of the released images and predicts a bounded residual. The same $\bar{z}$ follows a direct bypass to the final add-and-clip operation,

$$\hat{x}_m = \mathrm{clip}\left(\bar{z} + 0.5 \tanh D\left(\{a_\ell\}_{\ell=1}^{4}, \bar{z}\right), 0, 1\right). \tag{6}$$

Equation (6) supports any nonempty set size and preserves a direct observation-to-reconstruction path through the pixel mean. It also limits each output channel to $[\bar{z} - 0.5, \bar{z} + 0.5] \cap [0, 1]$. Both B_8 and $J(8)$ use $\bar{z}_8$ for this bypass and have the same output range; they differ in what is encoded before aggregation. Strong ISP color shifts can put the clean target outside this interval. We compare this head with a direct sigmoid output under paired training.

Training cycles through $m \in \{1, 2, 4, 8\}$ with the objective

$$\mathcal{L} = \mathcal{L}_1 + 0.5\mathcal{L}_{\mathrm{multi}} + 0.2\mathcal{L}_{\mathrm{grad}} + 0.05\mathcal{L}_{\mathrm{freq}} + 0.25\mathcal{L}_{\mathrm{id}}. \tag{7}$$

The first four terms in Eq. (7) compare pixels, mean-pooled images at 32 and 16 pixels, horizontal and vertical finite differences, and $\log(1 + |\mathcal{F}(x)|)$ for orthonormal real Fourier transforms. These terms use the 64-pixel target. The last term compares the upsampled prediction with the original 112-pixel target by frozen AdaFace cosine distance. ArcFace and AdaFace determine development selection; MagFace is excluded from attack fitting, model selection, and matched baseline selection. Its score therefore measures selection-independent transfer.

5 Experimental Protocol

Forward model and fitting. Aligned 112×112 RGB crops are divided by 255 and bilinearly resized to 64×64. Independent Gaussian noise with $\sigma = 0.01$ is added in this normalized RGB space, followed by clipping to $[0, 1]$. The released CamPro model applies $T(u) = g(\mathrm{clip}(Cu^{2.2}, 0, 1))$ per pixel, where C is its learned color matrix and g its 31-segment piecewise-linear gamma map. The attacked release contains no enhancer. Primary observations remain floating point; the PNG and JPEG experiments quantize and encode after the ISP. Reconstructed 64×64 images are upsampled to 112×112 for recognition. The perturbation is an RGB simulation, not calibrated sensor noise. The architecture, loss and 1,600-step budget were fixed in preliminary optical experiments. A numerical error was subsequently found in their high-order Zernike evaluation, and those optical results are excluded from this study. CamPro uses neither that optical basis nor those inverse weights. Its inverse is fitted separately through the stated ISP.

Data and model selection. FaceScrub supplies 180 training identities with eight images each and 32 development identities (Ng & Winkler, 2014). The test cohorts contain 96 FaceScrub and 32 NTU RGB+D 120 identities (Liu et al., 2020), each with one clean gallery image and four source queries. The NTU cohort receives no attack adaptation. Gallery and query images contain no exact or near duplicates; NTU gallery and query images also come from different source videos. The single-output pool contains seven CamPro models, including dedicated 1,600- and 6,000-step fits, identity-optimized and output-head variants, the primary set model and Burstormer. Selection maximizes the smaller ArcFace/AdaFace development Rank-1 at one input, breaking ties by PSNR and a fixed candidate order. It chooses the dedicated 1,600-step fit.

Primary fitting uses AdamW, batch size 12, learning rate 2×10^{-4}, weight decay 10^{-4}, cosine decay and gradient clipping at five. The joint model cycles through $m \in \{1, 2, 4, 8\}$ and uses the model after 1,600 CamPro updates under the fixed recipe, without a new eight-input search. The A_0 and B_8 fits share initialization, source order, architecture and loss, with one 1,600-update fit per condition and the terminal model retained. Comparisons reuse S_{e} and J with their original selection histories. Update counts match, but observation and encoding budgets differ. Evaluation uses nested prefixes of length 1, 2, 4, 8 and 16. ArcFace and AdaFace contribute to selection; MagFace is excluded from fitting and selection (Deng et al., 2019; Kim et al., 2022a; Meng et al., 2021).

Independent synthetic fits. DigiFace-1M supplies 1,000 fitting identities and separate 500-identity development and test groups (Bae et al., 2023). Each test identity has one gallery image and four queries, evaluated under five capture streams. Three seeds repeat fitting and selection. Per seed, four training stages from each of two single-output schedules and one set schedule form twelve

single-input candidates and four eight-input candidates. Long single-output and set schedules each encode 72,000 observations, but decode 72,000 and 19,200 targets respectively. This is an encoded-observation match, not equal decoder supervision or a convergence criterion. The single-output schedules use 1,600 or 6,000 updates; the set schedule uses 1,600. Each is evaluated at quarter-budget intervals on development identities.

Inference and statistics. Rates average four queries within each identity and then weight identities equally. Paired intervals resample identities, not individual queries or training seeds. Primary trajectories use 10,000 bootstrap draws. Selected attack contrasts use the empirical distribution of within-identity differences and Bonferroni correction. The human selected-attack comparisons retain their original 36-cell family, and synthetic comparisons retain six cells per contrast and seed, including optical comparisons omitted from this paper. The family sizes are not reduced when displaying only CamPro. The human cohorts are reused by the follow-up analyses; the three synthetic fits share their test identities and are reported separately. Input-matched comparisons use 100,000 identity resamples and an 18-cell family: $B_8 - A_0$, $J(8) - B_8$ and $A_0 - S_{\mathrm{e}}(1)$ across two cohorts and three recognizers.

Open-set tests use 3,072-identity galleries. Thresholds are calibrated either on 1,000 clean unknown identities or on their protected-and-reconstructed counterparts and tested on 1,000 disjoint unknowns. Unknown identities come from LFW (Huang et al., 2007). We report realized false-positive identification rate, FPIR, and target true-positive identification rate, TPIR, requiring both correct rank and threshold acceptance. Source and calibration roles remain separate throughout.

6 Results

6.1 Additional observations and ordinary fusion

Table 1 compares the independently selected single-output attack, image averaging and joint inversion. FaceScrub gains over the single attack range from 14.6 to 20.3 percentage points. The corresponding NTU gains range from 3.9 to 17.2 points. Only AdaFace's adjusted interval excludes zero. Joint processing also exceeds image averaging on FaceScrub by 8.9–15.4 points. All three adjusted intervals include or touch zero on NTU.

Table 1: CamPro with synthetic pre-ISP noise. S_{e} is selected at one input; Image 8 averages its eight independent reconstructions. Differences are percentage points with the original 36-comparison-adjusted identity-bootstrap intervals. Endpoints are Rank-1 proportions. Differences use unrounded means.

Cohort	Recognizer	$S_{\mathrm{e}}(1)$	Image 8	$J(8)$	$J(8) - S_{\mathrm{e}}(1)$	$J(8) -$ Image 8
FaceScrub	ArcFace	.362	.411	.565	20.3 [12.8, 28.1]	15.4 [8.9, 22.1]
	AdaFace	.445	.510	.615	16.9 [9.6, 24.5]	10.4 [3.4, 17.7]
	MagFace	.339	.396	.484	14.6 [7.0, 22.1]	8.9 [1.8, 15.9]
NTU	ArcFace	.188	.266	.289	10.2 [−2.3, 23.4]	2.3 [−7.8, 12.5]
	AdaFace	.141	.195	.312	17.2 [3.9, 32.0]	11.7 [0.0, 23.4]
	MagFace	.148	.148	.188	3.9 [−6.2, 14.1]	3.9 [−8.6, 14.1]

Independent selection also changes the synthetic comparison. In the first fit, the ArcFace gap falls from 16.5 points for the fixed set estimator to 7.8 points over the selected single-output attack, as shown in Table 2. Across three fits, mean gains range from 5.9 to 9.2 points across recognizers. These are three-fit means, not pooled confidence intervals. Each fit retains positive adjusted selected-attack gains. Relative to averaging the selected single-output reconstructions, joint inversion adds only 2.6, 3.0 and 1.7 points for ArcFace, AdaFace and MagFace in the first fit; MagFace's adjusted interval includes zero. Much of this synthetic recovery is therefore available through ordinary image fusion. Gallery size, training data and single-output strength differ from the human study, so the contrast does not isolate a human-versus-synthetic effect.

Table 2: First synthetic CamPro fit. S_{s} and J_{s} are selected separately; Image 8 averages S_{s} reconstructions. Rates average four queries and five captures within each of 500 identities. Gains and their unadjusted paired 95% intervals are in percentage points. After the original six-cell correction, all joint-minus-single intervals remain positive, whereas the MagFace joint-minus-image interval includes zero.

Recognizer	$S_{\mathrm{s}}(1)$	Image 8	$J_{\mathrm{s}}(8)$	$J_{\mathrm{s}}(8) - S_{\mathrm{s}}(1)$	$J_{\mathrm{s}}(8) -$ Image 8
ArcFace	0.281	0.333	0.359	7.8 [6.2,9.3]	2.6 [0.9, 4.3]
AdaFace	0.277	0.336	0.366	8.9 [7.3,10.5]	3.0 [1.3, 4.7]
MagFace	0.194	0.234	0.251	5.7 [4.5,7.0]	1.7 [0.3, 3.1]

6.2 Noiseless, pixel-mean and set inputs

Table 3 compares the input-matched A_0 and B_8 fits with S_{e} and J on the same human cohorts. The noisy methods receive the same first-eight observation sequence, with $S_{\mathrm{e}}(1)$ using its first member.

Table 3: Noiseless, pixel-mean and set inversion. Endpoints are Rank-1 proportions; $J(8) - B_8$ is in percentage points with 18-comparison-adjusted identity-bootstrap intervals. S_{e} and J reproduce Table 1. A_0 and B_8 are evaluated after 1,600 updates.

Cohort	Recognizer	A_0	$S_{\mathrm{e}}(1)$	B_8	$J(8)$	$J(8) - B_8$
FaceScrub	ArcFace	.341	.362	.406	.565	15.9 [9.1, 23.2]
	AdaFace	.406	.445	.518	.615	9.6 [3.1, 16.1]
	MagFace	.318	.339	.406	.484	7.8 [1.0, 14.8]
NTU	ArcFace	.156	.188	.172	.289	11.7 [0.0, 24.2]
	AdaFace	.211	.141	.188	.312	12.5 [3.1, 22.7]
	MagFace	.078	.148	.125	.188	6.2 [−4.7, 17.2]

On FaceScrub, $B_8 - A_0$ is 6.5–11.2 percentage points; adjusted intervals exclude zero for AdaFace and MagFace. Joint inversion exceeds B_8 by 7.8–15.9 points, with positive adjusted intervals for all three recognizers. On NTU, only AdaFace's adjusted $J(8) - B_8$ interval excludes zero. This ordering concerns the fitted models and budgets; it does not test whether the pixel mean is statistically sufficient for identity.

The pixel mean estimates $\mu_\sigma(x) = \mathbb{E}[z \mid x]$, which need not equal $T(x)$. Mean-domain recovery can therefore benefit from both reduced sampling variability and a different mean mapping. For encoder E, pooling after encoding also differs from encoding the pixel mean: $m^{-1}\sum_j E(z_j)$ generally differs from $E(\bar{z}_m)$. The comparison does not isolate feature variance as the source of the joint model's gain.

6.3 Cardinality and acquisition conditions

Figure 2 holds the joint model fixed. Eight observations raise Rank-1 from .333–.440 to .484–.615 on FaceScrub and from .094–.156 to .188–.312 on NTU. The sixteen-input results need not improve further for every recognizer. These trajectories describe one estimator, whereas Table 1 uses independent single-output selection.

In a separate common-execution sweep, varying only the pre-ISP perturbation scale gives exactly zero eight-versus-one gain at $\sigma = 0$. At $\sigma = 0.0025$, gains are 4.4–5.5 percentage points on FaceScrub and 0.8–3.9 points on NTU; at the training scale .01 they are 16.7–21.1 and 10.9–20.3 points. The fixed inverter makes this a sensitivity analysis, not a comparison of separately optimized attacks at every noise level. Exact frame copies leave feature moments and the pixel-mean bypass unchanged.

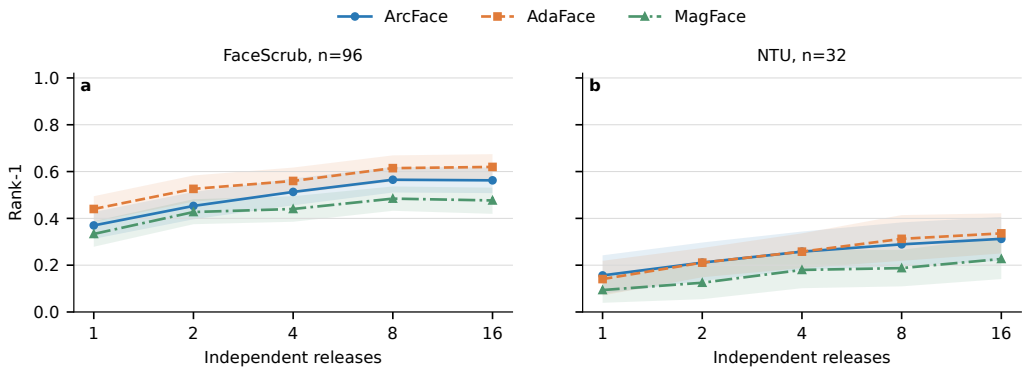


Figure 2: Fixed-estimator CamPro trajectories on nested observations. Every point uses the same model after 1,600 CamPro updates under the fixed training recipe. Attacks are not reselected at each release count. Shading denotes pointwise 95% identity-bootstrap intervals. Training uses at most eight inputs.

6.4 Deterministic collisions and noisy-output distinguishability

Eight of twelve constructed noiseless-equivalent color pairs have distinguishable noisy output distributions under the simultaneous CDF analysis, including three fully clipped pairs. The same eight pairs are detected in both sampling repetitions and both evaluated precisions.

Every column of the nonsingular color matrix C has negative sum, so at least one channel is lower-clipped for nonzero $u \geq 0$. The deterministic per-pixel Jacobian has rank at most two within each active branch. Nonzero gamma slopes preserve rank there; different branches can share outputs. Moving in the null space of the unclipped rows constructs four equivalent pairs per initial unclipped-channel count, before drawing noise. Table 4 shows the first completely clipped pair in construction order.

Table 4: Illustrative completely clipped pair. Both inputs map to (b_0, b_0, b_0) without noise, where $b_0 = .2679649591$. The CDF is evaluated at the selected red-channel threshold $t = b_0$; brackets are simultaneous 95% DKW ranges, rounded outward. Input coordinates are rounded for display.

	Input RGB	Count $Z_R \leq b_0$	Estimated CDF [range]
u_0	$(.754889, .534559, .381745)$	$65338/65536$	$.99698$ $[.988, 1.000]$
u_1	$(.760587, .535276, .354148)$	$60193/65536$	$.91847$ $[.909, .928]$

Each repetition uses 65,536 draws per color at $\sigma = .01$. For each pair we fix the first repetition's channel and threshold t with the largest empirical CDF distance. With $p_i = \Pr(Z_c \leq t \mid u_i)$, the count in m independent releases follows $\mathrm{Bin}(m, p_i)$. The statistic counts an output-threshold event; it does not directly observe the ISP clipping mask.

Under equal priors the ideal count classifier has accuracy

$$A_m^{\mathrm{count}} = \frac{1}{2} \sum_{k=0}^{m} \max\{\Pr_{\mathrm{Bin}(m,p_0)}(k), \Pr_{\mathrm{Bin}(m,p_1)}(k)\}. \tag{8}$$

For the displayed pair, substituting its CDF estimates gives accuracies .539, .575, .638 and .735 at $m = 1, 2, 4, 8$. With these probabilities, the eight-observation rule chooses u_0 only when all eight outputs satisfy the threshold event, and u_1 otherwise. The simultaneous DKW bands give an eight-release range $[.681, .766]$. Across all twelve pairs, the eight-release estimates range from .500 to .904. These are binomial-model calculations for constructed colors, not measured classification accuracies.

Write the random channel as $P_\sigma(\cdot \mid u)$. If it factored as $K_\sigma(\cdot \mid T(u))$ for an input-independent random postprocessor, equal noiseless outputs would imply equal noisy distributions. The observed

distribution differences provide evidence against this factorization: deterministic ambiguity need not survive pre-ISP perturbation. This does not order the two channels' informativeness for every task.

On 32 development identities, clipping-state changes rise from 26.4% to 41.3% as σ increases from .0025 to .02. Their within-scale correlations with fixed-J AdaFace gains range from $-.15$ to .00; all unadjusted 95% identity-bootstrap intervals span zero. This all-pixel switch count does not weight change magnitude or facial location and did not provide a clear predictor of identity recovery in this diagnostic.

6.5 Export and alignment

Export-domain adaptation retains 13.3–18.2-point gains on FaceScrub after 8-bit lossless PNG. JPEG95 point estimates are smaller, at 6.2–9.4 points. The direct paired JPEG-minus-PNG difference excludes zero after the original 12-comparison adjustment only for FaceScrub/ArcFace and NTU/MagFace. All NTU within-domain PNG and JPEG gain intervals include zero. These differences include both domain-specific attack selection and export transformation, rather than a codec-only effect.

Translation before protection exposes the importance of registration. Four-pixel shifts on the 64×64 grid reduce FaceScrub ArcFace eight-input Rank-1 to .055; blind phase-correlation registration restores it to .518. This controlled translation test is not a natural-video evaluation.

6.6 Restoration quality and identification

Higher identity recovery need not coincide with higher fidelity. In the selected synthetic CamPro comparison, eight-output PSNR is 20.28 dB versus 21.67 dB for the single-output model despite higher Rank-1. A direct sigmoid output head improves eight-input PSNR by .94 dB on FaceScrub and 2.29 dB on NTU; the six paired Rank-1 intervals include zero.

The mechanism-fitted Burstormer model has identical one- and eight-input correctness indicators on every CamPro query under the tested budget. CodeFormer and DiffBIR postprocessing lower the joint reconstruction's Rank-1 point estimates in all six CamPro cells. These are evaluations of the specified pretrained postprocessors, not mechanism-trained generative inverses.

6.7 Relative rank and threshold acceptance

With thresholds calibrated on reconstructed unknowns at nominal FPIR 1%, CamPro eight-release TPIR is only .3–1.6% on FaceScrub and zero on NTU, at realized FPIR .8–1.6%. The same 3,072-identity galleries and frozen clean-domain thresholds give genuinely clean target TPIR of 57.8–70.1% on FaceScrub and 6.2–21.9% on NTU. The protocol thus limits NTU acceptance even before protection. Unknowns are drawn from LFW, not the NTU acquisition domain. These results characterize the stated gallery and calibration population; they neither convert high Rank-1 into reliable attribution nor establish an upper bound on open-set attacks.

6.8 Native person-detection utility

We evaluate CamPro's released person detector on 1,000 previously used COCO development images, including images without people (Lin et al., 2014). The native path resizes the long side to 640 pixels, pads to a square, then applies normalized RGB perturbation and the same ISP. Outputs remain floating point without enhancement. Both protected conditions use the same fixed detector; the clean reference uses its corresponding released model. Person bbox AP averages IoU thresholds .50–.95 over the complete prediction set.

With the fixed CamPro detector, adding $\sigma = .01$ noise lowers person AP from .518 to .505, a change of -1.34 percentage points with paired-image 95% interval $[-1.88, -.77]$. The 1,000 bootstrap resamples condition on one fixed noise realization per image. Detection uses the native full-frame path, whereas the identity experiments apply the ISP after resizing face crops to 64 pixels.

Table 5: Fixed-weight native detection. All entries are person bbox AP proportions. S, M and L denote COCO object-area groups. The clean model differs from the protected-domain detector; only the last two rows isolate added noise.

Input	AP	AP50	AP75	AP_S	AP_M	AP_L
Clean reference	.624	.873	.680	.412	.685	.806
CamPro, $\sigma = 0$	.518	.782	.534	.266	.589	.748
CamPro, $\sigma = .01$	.505	.759	.535	.254	.578	.735

7 Discussion and Conclusion

Repeated CamPro measurements improved closed-gallery recovery under the specified synthetic conditions. The additional value of joint inversion varied across settings; ordinary image averaging recovered much of the synthetic gain. On FaceScrub, joint inversion also outperformed the inverse trained on eight-output pixel means. At native detection resolution, the same ISP and perturbation definition retained person AP .505, compared with .518 without the added noise.

Pre-ISP perturbations distinguish some colors that the noiseless CamPro mapping merges. For the constructed binary problems, calculated count discrimination improves with the number of observations and varies across pairs. This provides a possible channel-level source of additional distinguishability without quantifying its contribution to the identity gains.

The attack comparisons remain budget-defined, with one fit per new input condition. Task and identity experiments use different spatial preprocessing paths, so the reported detection and recovery rates do not describe one shared full-frame release.

Ethics Statement. The study uses existing research datasets without new participant interaction. Identity-bearing images, names, gallery templates and reconstructions are not redistributed. Dataset access and use remain subject to the original terms. The study does not seek participants' real names.

Reproducibility Statement. The experiments use fixed identity partitions, common observation sequences and development-only model selection. The selected synthetic inverse weights are available for inference, but the primary human-cohort inverse weights are not publicly distributed. Statistical recomputation and independent retraining therefore have different reproduction requirements.

AI Use Statement. AI assistance supported manuscript polishing, simulated peer review and verification. The authors are responsible for the final manuscript and reported research.

Conflicts of Interest. The authors declare no conflicts of interest.

References

Miika Aittala and Frédo Durand. Burst image deblurring using permutation invariant convolutional neural networks. In *Computer Vision – ECCV 2018*, pp. 748–764. Springer, 2018. doi: 10.1007/978-3-030-01237-3_45.

Gwangbin Bae, Martin de La Gorce, Tadas Baltrušaitis, Charlie Hewitt, Dong Chen, Julien Valentin, Roberto Cipolla, and Jingjing Shen. DigiFace-1M: 1 million digital face images for face recognition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3526–3535, 2023.

Hao Chen, Pramod K. Varshney, and J. H. Michels. Noise enhanced parameter estimation. *IEEE Transactions on Signal Processing*, 56(10):5074–5081, 2008. doi: 10.1109/TSP.2008.928508.

Jiacheng Cheng, Xiang Dai, Jia Wan, Nick Antipa, and Nuno Vasconcelos. Learning a dynamic privacy-preserving camera robust to inversion attacks. In *Computer Vision – ECCV 2024*, Lecture Notes in Computer Science, pp. 349–367. Springer, 2024. doi: 10.1007/978-3-031-72897-6_20.

Mauricio Delbracio and Guillermo Sapiro. Burst deblurring: Removing camera shake through fourier burst accumulation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2385–2393, 2015. doi: 10.1109/CVPR.2015.7298852.

Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4690–4699, 2019. doi: 10.1109/CVPR.2019.00482.

Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burst image restoration and enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5759–5768, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Dudhane_Burst_Image_Restoration_and_Enhancement_CVPR_2022_paper.html.

Akshay Dudhane, Syed Waqas Zamir, Salman Khan, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Burstormer: Burst image restoration and enhancement transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5703–5712, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Dudhane_Burstormer_Burst_Image_Restoration_and_Enhancement_Transformer_CVPR_2023_paper.html.

Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical Poissonian–Gaussian noise modeling and fitting for single-image raw-data. *IEEE Transactions on Image Processing*, 17(10):1737–1754, 2008. doi: 10.1109/TIP.2008.2001399.

Tal Hassner, Iacopo Masi, Jungyeon Kim, Jongmoo Choi, Shai Harel, Prem Natarajan, and Gérard Medioni. Pooling faces: Template based face recognition with pooled face images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 59–67, 2016. doi: 10.1109/CVPRW.2016.23.

Carlos Hinojosa, Juan Carlos Niebles, and Henry Arguello. Learning privacy-preserving optics for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2573–2582, 2021. doi: 10.1109/ICCV48922.2021.00257.

Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts Amherst, 2007.

Bhavin Jawade, Deen Dayal Mohan, Dennis Fedorishin, Srirangaraj Setlur, and Venu Govindaraju. CoNAN: Conditional neural aggregation network for unconstrained face feature fusion. In *2023 IEEE International Joint Conference on Biometrics*, pp. 1–10, 2023. doi: 10.1109/IJCB57857.2023.10448797.

Minchul Kim, Anil K. Jain, and Xiaoming Liu. AdaFace: Quality adaptive margin for face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18750–18759, 2022a. doi: 10.1109/CVPR52688.2022.01819.

Minchul Kim, Feng Liu, Anil K. Jain, and Xiaoming Liu. Cluster and aggregate: Face recognition with large probe set. In *Advances in Neural Information Processing Systems*, volume 35, pp. 36054–36066, 2022b. doi: 10.52202/068431-2613.

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 *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3744–3753. PMLR, 2019.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *Computer Vision – ECCV 2014*, pp. 740–755, 2014. doi: 10.1007/978-3-319-10602-1_48.

Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. DiffBIR: Toward blind image restoration with generative diffusion prior. In *Computer Vision – ECCV 2024*, pp. 430–448. Springer, 2024. doi: 10.1007/978-3-031-73202-7_25.

Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C. Kot. NTU RGB+D 120: A large-scale benchmark for 3d human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10):2684–2701, 2020. doi: 10.1109/TPAMI.2019.2916873.

Yu Liu, Junjie Yan, and Wanli Ouyang. Quality aware network for set to set recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5790–5799, 2017. doi: 10.1109/CVPR.2017.499.

Jhon Lopez, Carlos Hinojosa, Henry Arguello, and Bernard Ghanem. Privacy-preserving optics for enhancing protection in face de-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12120–12129, 2024. doi: 10.1109/CVPR52733.2024.01152.

Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. MagFace: A universal representation for face recognition and quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14225–14234, 2021. doi: 10.1109/CVPR46437.2021.01400.

Hong-Wei Ng and Stefan Winkler. A data-driven approach to cleaning large face datasets. In *IEEE International Conference on Image Processing*, pp. 343–347, 2014. doi: 10.1109/ICIP.2014.7025068.

Seong Joon Oh, Rodrigo Benenson, Mario Fritz, and Bernt Schiele. Faceless person recognition: Privacy implications in social media. In *Computer Vision – ECCV 2016*, volume 9907 of *Lecture Notes in Computer Science*, pp. 19–35. Springer, 2016. doi: 10.1007/978-3-319-46487-9_2.

Francesco Pittaluga and Sanjeev J. Koppal. Privacy preserving optics for miniature vision sensors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 314–324, 2015. doi: 10.1109/CVPR.2015.7298628.

Fan Wei and Richard M. Dudley. Two-sample Dvoretzky–Kiefer–Wolfowitz inequalities. *Statistics & Probability Letters*, 82(3):636–644, 2012. doi: 10.1016/j.spl.2011.11.012.

Jiaolong Yang, Peiran Ren, Dongqing Zhang, Dong Chen, Fang Wen, Hongdong Li, and Gang Hua. Neural aggregation network for video face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4362–4371, 2017. doi: 10.1109/CVPR.2017.554.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Shangchen Zhou, Kelvin C. K. Chan, Chongyi Li, and Chen Change Loy. Towards robust blind face restoration with codebook lookup transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 30599–30611, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/c573258c38d0a3919d8c1364053c45df-Abstract-Conference.html.

Wenjun Zhu, Yuan Sun, Jiani Liu, Yushi Cheng, Xiaoyu Ji, and Wenyuan Xu. CamPro: Camera-based anti-facial recognition. In *Network and Distributed System Security Symposium*, 2024. doi: 10.14722/ndss.2024.24158.

APPENDIX

A CAMPRO IMPLEMENTATION AND DATA

A.1 RELEASE AND RECOGNITION INTERFACES

The released CamPro model contains a color correction matrix C and gamma knots $b_0, \ldots, b_K$. For normalized RGB input u, the implemented mapping is

$$v = \mathrm{clip}(Cu^{2.2}, 0, 1), \qquad T(u)_c = (1 - \alpha_c) b_{k_c} + \alpha_c b_{k_c+1}, \tag{9}$$

where $k_c = \lfloor K \,\mathrm{clip}(v_c, 10^{-6}, 1 - 10^{-6}) \rfloor$ and $\alpha_c = K v_c - k_c$. The matrix acts on color channels at each pixel; the interpolation is channelwise. Input clipping follows the added noise, and another clip follows color correction. The index clamp selects a valid gamma segment, including at the endpoints. The primary release uses this floating-point mapping without the optional enhancer or digital encoding. Clipping can merge distinct inputs, so knowledge of T alone does not imply a global inverse.

The tested mapping is `CamPro-b75e73e1087f`, the released default model at source revision `2aeac7b832e9`. Its color matrix is

$$C \simeq \begin{pmatrix} .431650 & -.733373 & -.649450 \\ -.413554 & .773158 & -.315060 \\ -.294347 & -.323411 & .754818 \end{pmatrix}.$$

There are $K = 31$ gamma segments. The PDF attachment `campro-configuration.json` gives the full-precision matrix and 32 knots, full source revision, model and recognizer SHA-256 identifiers, and the seven candidate configurations C1–C7. It contains no face data or human inverse weights. The source is the public CamPro repository.

Proposition (local rank of the deterministic color map). For $u \in (0,1)^3$ within a fixed clipping region and away from active gamma knots, $\mathrm{rank}\, J_T(u) \le 2$. *Proof.* The column sums of C are $(-.276251, -.283626, -.209692)$, so $\mathbf{1}^\top C u^{2.2} < 0$. At least one channel is lower-clipped. With M the diagonal unclipped-channel mask,

$$J_T(u) = D_g \, M \, C \, \mathrm{diag}(2.2 u^{1.2}), \qquad \mathrm{rank}\, J_T(u) \le \mathrm{rank}\, M \le 2.$$

The nonsingular matrix and nonzero gamma slopes preserve rank on each active branch. The inactive directions depend on u; this is not a bound on the information in $p_\sigma(z \mid u)$. The maximum preclip channel values are $(.431650, .773158, .754818)$, so upper clipping is inactive. The largest reachable gamma indices are 13, 23 and 23. Their respective reachable slope counts are 9/5, 15/9 and 15/9 (positive/negative), compared with 19/12 over the entire 31-segment curve.

The resampling and perturbation order follows Section 5, with bilinear interpolation not aligning corner centers. AdaFace uses RGB normalized to $[-1, 1]$, MagFace uses BGR in $[0, 1]$, and the ArcFace ONNX interface uses its uint8 color conversion and 127.5 centering and scaling. Gallery and query paths use the same recognizer-specific interface. Clean open-set targets use the original 112-pixel crops without protection or inversion.

A.2 Source and identity separation

Training and development identities are disjoint from both test cohorts. Each query generates its own release set. NTU gallery and query images come from different source videos, without adjacent frame pairs; 30 of its 32 identities also span four query actions. FaceScrub checks found no exact or near-duplicate gallery-query pairs; the largest pairwise PSNR was 19.38 dB. These checks do not address unknown overlap in recognizer pretraining. Human-cohort follow-ups reuse the same test identities and are not independent replications.

DigiFace uses the first 2,000 selected identities, split into 1,000 training, 500 development and 500 test identities. All 272 identities used in the earlier synthetic pilot remain in the training group. The eight fitting views are 0, 9, 18, 27, 36, 45, 54 and 63; development and test use view 0 for the gallery and views 18, 36, 54 and 71 for queries. Rendered images are resized from 112 to 64 pixels without face alignment. Five independently keyed 16-member sequences are generated for each query. Keys include identity, view, capture repetition, release position and random-source type. Prefix length and inference batch size do not change a member. The single-output attack always receives the first member, with no best-of-eight selection.

Implementation details. The architecture in Section 4 has 3,305,059 parameters. Moment reductions return from float64 to the model dtype. AdamW's cosine schedule ends at 10^{-5}; training never uses cardinality 16.

Native detector evaluation. The 1,000-image COCO val2017 development subset contains 2,072 person annotations, including 41 crowd regions; 520 images contain a person. The image list is unchanged from earlier task development. Both person detectors use the released single-class YOLOv5m

architecture and fixed weights. The official clean model has SHA-256 prefix `ead347a4d78e`; the CamPro model has prefix `7b0f5811be31`. Long-side resizing uses area interpolation when reducing and linear interpolation when enlarging, followed by square padding at RGB value 114/255. Perturbation follows padding. NMS uses confidence .001, IoU .6 and at most 300 boxes; official COCO accumulation uses at most 100 per image. Source-coordinate boxes use the recorded resize and padding. One per-image keyed noise draw, seed 20260921, is shared by the saved prediction and resampling analyses. The paired bootstrap resamples whole image records and rebuilds dataset precision–recall curves, including crowd handling and images without people. There is no detector fitting or new utility threshold. The PDF attachment `campro-utility-and-input-comparison.json` contains full model identifiers, image IDs, endpoint statistics and paired intervals.

B Attack Selection and Repeated Fits

J is the terminal 1,600-step human CamPro set fit of the inherited recipe. S_{d} selects a dedicated one-output fit from the 1,600- and 6,000-step schedules. S_{e} selects from the expanded pool in Table 6. Both CamPro pools choose the dedicated 1,600-step model. One-output development selection maximizes the smaller ArcFace/AdaFace Rank-1, then PSNR, then the fixed candidate order. Burstormer's one-input candidate repeats one unique measurement in its eight input positions. MagFace is not used in fitting or selection.

C1–C3 are random-initialized primary-loss fits: 1,600 single-output, 6,000 single-output and 1,600 set updates, respectively. C4 starts from C3 and takes 1,600 further updates with $L_{\mathrm{Ada}} + .1L_{\mathrm{CE}} + .05L_1$. Here L_{CE} contrasts the 12 predicted AdaFace embeddings against the corresponding clean targets at temperature .07; target embeddings use 112-to-64-to-112 resampling. AdamW uses 10^{-4} decaying to 5×10^{-6}, batch size 12 and the 1/2/4/8 cycle. C5 and C6 are random-initialized bounded/direct-head fits, each trained for 1,600 updates with the primary loss; eight-input development selection retains steps 1,200 and 1,600. C7 is random-initialized Burstormer, trained for 9,000 eight-input updates at batch size one; step 9,000 is retained. All seven retained models are then compared at one unique observation.

Table 6: Human one-observation development selection on 32 identities, four queries each. PSNR uses the 64-pixel grid. The dedicated 1,600-step fit is selected by the smaller ArcFace/AdaFace score; MagFace and test scores do not enter selection.

ID	Candidate	ArcFace	AdaFace	PSNR	Selected
C1	Dedicated 1,600	0.523	0.641	18.18	Yes
C2	Dedicated 6,000	0.477	0.547	19.22	No
C3	Primary set	0.516	0.555	18.02	No
C4	Identity-optimized set	0.492	0.562	16.93	No
C5	Paired bounded head	0.508	0.562	17.79	No
C6	Paired direct head	0.500	0.578	18.59	No
C7	Burstormer	0.289	0.492	16.08	No

Image fusion averages independently reconstructed RGB images; feature fusion averages and renormalizes embeddings. Inversion of the protected pixel mean is a separate comparison. Human contrasts share the first eight members of the primary 16-member sequence.

Matched-input fits. A_0 and B_8 use the same 180 auxiliary training identities and seed 20261210 for initialization and source scheduling. Both follow the primary loss, batch size and 1,600-step schedule, retaining the terminal model. The pixel mean is accumulated in float64 and converted to float32 before the unmodified one-member encoder. The original one-member variance stabilizer is retained during both fitting and inference. The new models are fixed before historical-cohort evaluation; development scores do not select or extend these fits. MagFace is used only after fitting. A_0, S_{e}, B_8 and J process 19,200, 19,200, 153,600 and 72,000 training observations, respectively, but B_8 encodes only 19,200 mean images. Update matching does not equalize encoder work or establish convergence. Training preserves the existing CUDA pooling backward operation, which is not bitwise deterministic; inference uses deterministic algorithms with TF32 disabled.

The new comparisons resample complete identities 100,000 times, with an 18-cell family covering three contrasts, two cohorts and three recognizers. Table 7 complements the main joint-minus-mean comparison. All six adjusted $A_0 - S_e(1)$ intervals span zero.

Table 7: Remaining matched-input contrasts, in percentage points with 18-comparison-adjusted identity-bootstrap intervals. These condition on the fitted models and one replayed capture sequence.

Cohort	Recognizer	$B_8 - A_0$	$A_0 - S_e(1)$
FaceScrub	ArcFace	6.5 $[-1.0, 13.8]$	-2.1 $[-9.4, 4.9]$
	AdaFace	11.2 $[3.9, 18.2]$	-3.9 $[-11.2, 3.4]$
	MagFace	8.9 $[2.1, 15.6]$	-2.1 $[-8.1, 4.2]$
NTU	ArcFace	1.6 $[-9.4, 10.9]$	-3.1 $[-14.1, 7.8]$
	AdaFace	-2.3 $[-11.7, 7.8]$	7.0 $[-3.9, 18.0]$
	MagFace	4.7 $[-6.2, 17.2]$	-7.0 $[-16.4, 1.6]$

B.1 Synthetic selection and replication

Each seed fits dedicated one-output schedules of 1,600 and 6,000 updates and a 1,600-update set schedule, all at batch size 12. Candidate training stages are 400, 800, 1,200 and 1,600 for short schedules, and 1,500, 3,000, 4,500 and 6,000 for the long schedule. The set candidates are also evaluated at one input. Thus twelve candidates compete for S_s and four for J_s under the same two-recognizer development rule, followed by PSNR and candidate order. All three CamPro seeds choose the 6,000-step single fit and 1,600-step set fit.

The long single and set schedules each encode 72,000 observations, but their decoder supervision totals are 72,000 and 19,200 targets. The comparison is budget-defined. In the first long CamPro fit, the four development scores are .197, .264, .288 and .297, so the last evaluated stage is still improving. Table 9 reports every seed without pooling them. Table 8 separates capture variation from training variation, and Table 10 retains the fixed-model trajectory and clean/raw references. The first-fit single and set PSNRs are 21.67 and 20.28 dB respectively. Image-fusion endpoints appear directly in Table 2.

Table 8: First-fit selected-attack gains in percentage points for five capture streams. The final column is the original six-cell-adjusted interval for their mean.

Recognizer	Capture 1	2	3	4	5	Adjusted interval
ArcFace	6.9	8.1	8.5	6.6	8.8	[5.6, 9.9]
AdaFace	7.8	9.0	8.4	10.1	9.2	[6.7, 11.0]
MagFace	5.6	5.6	6.3	5.1	6.1	[4.0, 7.5]

Table 9: Three independently fitted and selected synthetic CamPro attacks. Differences and paired 95% intervals are in percentage points, computed from unrounded decisions. Seeds share identities and captures and are not pooled.

Seed	Recognizer	$S_s(1)$	$J_s(8)$	Difference [95% CI]
1	ArcFace	0.281	0.359	7.8 [6.2, 9.3]
1	AdaFace	0.277	0.366	8.9 [7.3, 10.5]
1	MagFace	0.194	0.251	5.7 [4.5, 7.0]
2	ArcFace	0.282	0.347	6.5 [5.0, 8.0]
2	AdaFace	0.269	0.355	8.7 [7.1, 10.2]
2	MagFace	0.190	0.253	6.3 [4.9, 7.6]
3	ArcFace	0.273	0.352	8.0 [6.5, 9.5]
3	AdaFace	0.266	0.367	10.1 [8.6, 11.6]
3	MagFace	0.190	0.247	5.7 [4.3, 7.0]

Table 10: Synthetic reference endpoints and first-fit fixed-J_s trajectory. Clean is clean-query recognition and Raw 1 uses the first protected member. All other columns use one eight-input-selected model.

Recognizer	Clean	Raw 1	1	2	4	8	16
ArcFace	.851	.018	0.194	0.271	0.326	0.359	0.380
AdaFace	.863	.053	0.196	0.282	0.336	0.366	0.388
MagFace	.744	.003	0.141	0.198	0.232	0.251	0.260

C Paired Statistical Inference

Each human identity contributes its four-query mean. For a fixed recognizer, let d_i be the within-identity difference between two attacks. We estimate $n^{-1}\sum_i d_i$ and resample complete identities with replacement. Primary curves use 10,000 percentile-bootstrap draws. The synthetic analysis first averages all four queries and five captures within each of 500 identities. Captures, queries and fitting seeds do not increase the independent identity count. All intervals condition on the fitted models, observed captures and, where applicable, fixed calibration thresholds.

For the retained human paired counts, $4d_i \in \{-4, \ldots, 4\}$. If c_k identities have value k, the coefficients of

$$\left[\sum_{k=-4}^{4} \frac{c_k}{n} t^{k+4}\right]^n \tag{10}$$

give the empirical bootstrap distribution, with coefficient index s corresponding to $(s-4n)/(4n)$. The synthetic analysis uses 41 bins and denominator 20. A between-export difference of gains uses 17 bins and denominator 4. Percentiles are the first values whose cumulative probability reaches the specified level. This deterministically recomputes the empirical resampling distribution; it does not establish exact population coverage.

Multiplicity families retain their original sizes, including optical cells even when only CamPro is displayed. The human expanded-pool comparison has 36 cells, comprising three contrasts across two mechanisms, two cohorts and three recognizers. Synthetic contrasts each retain a six-cell family per seed. Export adaptation has 72 within-domain comparisons and a separate 12-cell family for differences of gains. Set Transformer gains retain twelve cells. The earlier primary sensitivity analysis used 200,000 bootstrap draws with Bonferroni intervals and one million paired sign flips with Holm adjustment. Displayed unadjusted 95% intervals are labeled separately from family-adjusted intervals. All adjustments use the comparison families defined above.

When every observed identity has zero paired difference, empirical resampling is degenerate at zero. Likewise, no observed accepted target or deteriorating identity gives no empirical information about a nonzero population upper bound. Such cells are described by their event counts or marked as degenerate, not interpreted as equivalence or certainty about unseen identities. Unknown probe FPIR intervals use exact binomial intervals; target TPIR intervals cluster the four queries by identity. Calibration uncertainty is not included in intervals conditional on the selected threshold.

D Noise, Export and Alignment

D.1 Development switching diagnostic

Table 11 uses the existing 32 development identities, four queries each, with common standard-normal perturbations at all five scales. Clip-change is the fraction of pixel-channel positions taking both clipped and unclipped states among eight releases; segment-change similarly counts changes in the gamma segment index. All pixels, including zero-valued pixels, are included; the indicator does not weight change magnitude or facial location. Clip rates average sixteen releases. The same J and AdaFace evaluate one/eight prefixes in a 32-identity gallery. Correlations relate four-query identity means, with unadjusted 95% intervals from 10,000 whole-identity resamples. This is a development sensitivity analysis with one fixed noise-trained attack, not a matched-attack comparison at each scale.

Table 11: Fixed-J development diagnostic. Clip, Switch and Segment are percentages defined in the text. Endpoints are AdaFace Rank-1; gains are percentage points. ρ relates identity-mean Switch to the gain, with an unadjusted 95% interval; both variables are constant at zero noise.

σ	Clip	Switch	Segment	$J(1)$	$J(8)$	Gain	ρ [95% CI]
0	86.7	0.0	0.0	0.586	0.586	0.0	–
0.0025	79.0	26.4	1.4	0.570	0.617	4.7	-0.07 $[-0.44, 0.33]$
0.005	79.0	29.0	2.6	0.570	0.633	6.2	0.00 $[-0.40, 0.42]$
0.01	78.9	33.5	4.7	0.594	0.766	17.2	-0.08 $[-0.42, 0.28]$
0.02	78.5	41.3	7.6	0.539	0.727	18.8	-0.15 $[-0.53, 0.25]$

D.2 Deterministically equivalent colors

We select four pairs for each unclipped-channel count: zero, one and two. For $q_0 = u_0^{2.2}$ and the unclipped rows C_A, choose a unit vector $n \in \ker C_A$ and set $q_1 = q_0 + tn$. The positive step is half the distance to the first clipping-region (with 10^{-6} margin) or coordinate boundary, restricting $q \in [.05^{2.2}, .95^{2.2}]^3$. The first four valid pairs in each stratum from seed 20260921 are retained before noisy sampling, with $\|u_1 - u_0\|_2 \geq .01$. Noiseless numerical output discrepancies are below 9×10^{-16} in float64 and 1.41×10^{-6} in float32.

Each color receives $N = 65{,}536$ independent Gaussian RGB perturbations at $\sigma = .01$, followed by input clipping and the same ISP, without export. Domain-separated noise streams in two repetitions also compare u_0 with another sample from itself. For each RGB channel let $\widehat{d} = \sup_z |\widehat{F}_0(z) - \widehat{F}_1(z)|$. Combining one-sample DKW bands (Wei & Dudley, 2012) with a union bound gives simultaneous 95% intervals $[\max(0, \widehat{d} - b), \min(1, \widehat{d} + b)]$, where $b = 2\sqrt{\log(4M/.05)/(2N)} = .017509$ and $M = 12 \times 2 \times 3 \times 2 \times 2 = 288$ covers pairs, repetitions, channels, pair/control contrasts and float64/float32 evaluations. This bound allows atoms from clipping; no continuous-distribution KS approximation is used.

Positive lower bounds occur in both repetitions for 3/4, 2/4 and 3/4 pairs with zero, one and two unclipped channels. The largest same-color control distance is .00949; none has a positive lower bound. All pairs and repetitions are retained in `campro-color-channel.json`, embedded in the PDF and provided as a sidecar.

Finite-count analysis from saved CDF witnesses. For each pair, we select the channel with largest empirical CDF distance in the first float64 repetition (seed 20260921), breaking ties by RGB order. Its saved maximizing threshold is fixed for all $m \in \{1, 2, 4, 8\}$. Counts at this threshold give $\hat{p}_i = n_i/N$ without new sampling. The two colors have equal priors; Eq. 8 is the ideal classifier restricted to these counts, assuming the probabilities are known. The 65,536 draws estimate p_i; they are not the count classifier's observation budget. Witnesses were selected separately in the two repetitions; the second is not used as validation at the first repetition's threshold.

Let $I_i = [l_i, h_i] = [\max(0, \hat{p}_i - b/2), \min(1, \hat{p}_i + b/2)]$. The original DKW event covers all thresholds and channels, including the selected witnesses. Mapping $I_0 \times I_1$ through Eq. 8 therefore gives simultaneous 95% ranges for all twelve pairs and all four counts, without another comparison adjustment. For $p_0 \leq p_1$, binomial likelihood ratios are monotone in the count, so their total variation equals the largest difference between their CDFs. Each CDF decreases with its probability parameter; hence accuracy increases as the ordered probabilities move apart. The minimum over the rectangle is $1/2$ when the intervals overlap, otherwise the value at the nearest endpoints. The maximum is the larger of $A_m^{\text{count}}(l_0, h_1)$ and $A_m^{\text{count}}(h_0, l_1)$.

Table 12 retains all twelve pairs, including near-chance results. Pair 8 is the first completely clipped pair in construction order and supplies Table 4. Only witness counts, not full CDF curves, were saved. The PDF attachment and sidecar `campro-finite-count.json` contain all selected channels, thresholds, probability intervals and accuracy ranges. No identity data enter this analysis.

Table 12: Finite-count color discrimination for every constructed pair, in original order. a counts initially unclipped channels. Values substitute saved CDF probabilities in Eq. 8; brackets give simultaneous 95% ranges, rounded outward. These are ideal count-model accuracies, not measured attack accuracies.

Pair	a	$m = 1$	$m = 2$	$m = 4$	$m = 8$	Range at $m = 8$
0	1	.556	.605	.686	.796	$[.747, .827]$
1	1	.507	.511	.515	.521	$[.500, .548]$
2	1	.534	.558	.584	.607	$[.580, .634]$
3	2	.523	.536	.543	.563	$[.539, .587]$
4	1	.505	.508	.512	.515	$[.500, .543]$
5	2	.620	.698	.773	.833	$[.814, .852]$
6	2	.506	.509	.510	.515	$[.500, .538]$
7	2	.566	.610	.656	.700	$[.674, .724]$
8	0	.539	.575	.638	.735	$[.681, .766]$
9	0	.500	.500	.500	.500	$[.500, .535]$
10	0	.569	.625	.710	.802	$[.760, .844]$
11	0	.616	.702	.814	.904	$[.867, .943]$

D.3 Noise scale and alternative acquisition

Table 13 changes the perturbation scale while keeping the inverse fixed. The zero-noise condition gives identical measurements, and exact copies leave the architecture's moments and pixel mean unchanged. It is not a separately optimized zero-noise attack. Each intervention uses common random keys internally; the independent intervention sweeps are not the same realization as the primary sequence.

The two zero-noise FaceScrub/MagFace entries retain different recorded counts: 131/384 in Table 13 and 132/384 in Table 14. The source tensor, ISP and inverse-model identifiers agree. The executions differ in inverse batch size (16 versus 8), recognizer batch size (64 versus 32), and numerical settings. The earlier sweep retained aggregates but not query decisions or scores, so the changed query and exact cause cannot be recovered. Noise keys do not explain this zero-noise difference; neither recorded endpoint is replaced.

Table 13: Noise-scale sensitivity of one frozen attack in a separate common-execution sweep. Triples are ArcFace/AdaFace/MagFace Rank-1. No attack is reselected at a new noise scale.

Cohort	σ	$R(1)$	$R(8)$
FaceScrub	0	0.365/0.440/0.341	0.365/0.440/0.341
	0.0025	0.367/0.448/0.339	0.417/0.492/0.393
	0.0050	0.352/0.435/0.339	0.464/0.534/0.440
	0.0100	0.341/0.424/0.336	0.552/0.617/0.503
	0.0200	0.312/0.406/0.292	0.581/0.630/0.549
NTU	0	0.148/0.148/0.102	0.148/0.148/0.102
	0.0025	0.172/0.141/0.109	0.188/0.180/0.117
	0.0050	0.148/0.164/0.109	0.219/0.219/0.109
	0.0100	0.148/0.125/0.070	0.352/0.328/0.180
	0.0200	0.141/0.148/0.055	0.383/0.430/0.250

The additional scenarios in Table 14 include $v = aP + \sqrt{b}\,\xi$, with $P \sim \mathrm{Poisson}(x/a)$ and $\xi \sim \mathcal{N}(0, I)$. Before clipping, the conditional variance is $ax + b$. The low and high settings use $(a, b) = (10^{-4}, 10^{-6})$ and $(10^{-3}, 10^{-5})$. These signal-dependent scenarios follow a standard noise-model form (Foi et al., 2008), but their parameters are not fitted to hardware. A 3×3 spatial mean with zero padding is applied before the ISP in the averaging control; it is not a hardware-denoiser model. JPEG95 acts after the ISP. No scenario-specific retraining is performed in this table.

Table 14: Static-acquisition scenarios with the same frozen attack. Triples are ArcFace/AdaFace/MagFace. PG denotes the two specified Poisson–Gaussian settings; neither is sensor-calibrated. These are sensitivity tests, not matched retraining.

Cohort	Acquisition	$R(1)$	$R(8)$
FaceScrub	No noise	0.365/0.440/0.344	0.365/0.440/0.344
FaceScrub	Gaussian 0.0025	0.354/0.430/0.344	0.414/0.482/0.391
FaceScrub	Gaussian 0.01	0.375/0.435/0.357	0.552/0.607/0.482
FaceScrub	Poisson–Gaussian low	0.372/0.448/0.344	0.521/0.591/0.484
FaceScrub	Poisson–Gaussian high	0.328/0.398/0.281	0.625/0.667/0.576
FaceScrub	Gaussian + spatial mean	0.154/0.182/0.143	0.206/0.255/0.174
FaceScrub	Gaussian + JPEG95	0.029/0.039/0.026	0.042/0.042/0.029
NTU	No noise	0.148/0.148/0.102	0.148/0.148/0.102
NTU	Gaussian 0.0025	0.141/0.148/0.109	0.188/0.180/0.086
NTU	Gaussian 0.01	0.141/0.141/0.094	0.266/0.336/0.180
NTU	Poisson–Gaussian low	0.164/0.164/0.117	0.312/0.258/0.117
NTU	Poisson–Gaussian high	0.117/0.125/0.086	0.453/0.500/0.219
NTU	Gaussian + spatial mean	0.102/0.117/0.070	0.164/0.164/0.125
NTU	Gaussian + JPEG95	0.086/0.062/0.031	0.094/0.086/0.039

D.4 Quantization and export adaptation

Exactly the same $\sigma = .01$ protected observations are exported as float32, 8-bit PNG or JPEG95. PNG clamps RGB to $[0, 1]$, rounds to uint8, encodes and decodes losslessly, and divides by 255; decoded pixels equal the quantized array. JPEG95 starts from that uint8 array and uses OpenCV's default chroma subsampling. Its difference from PNG therefore includes color conversion, subsampling and transform quantization.

Domain adaptation starts from S_{d} and J. Both branches execute 1,600 AdamW updates at batch size 12, learning rate 10^{-4} decaying to 5×10^{-6}, weight decay 10^{-4} and gradient clipping at five. The primary loss and paired source/perturbation streams are retained. The single branch uses one input and the set branch cycles through 1, 2, 4 and 8, giving equal target counts but 19,200 versus 72,000 encoded observations. Initialization and steps 400, 800, 1,200 and 1,600 form each development pool. The independently selected single/set steps are 0/1,200 for float32, 800/400 for PNG and 1,600/1,200 for JPEG95. We denote the domain-selected attacks by $S_{\mathrm{x}}, J_{\mathrm{x}}$.

Tables 15 and 16 retain frozen and adapted endpoints, including the weak NTU results. For each identity and export domain d, let $g_{i,d}$ be the four-query joint-minus-single mean. Table 17 compares $g_{i,d_2} - g_{i,d_1}$ within identity. Source images, galleries and pre-export observations are shared, while attacks are selected within each domain. The comparison measures their combined change, rather than a codec-only effect.

Table 15: FaceScrub export endpoints. Frozen $J(8)$ is shown alongside domain-selected $S_{\mathrm{x}}(1)$, image averaging and $J_{\mathrm{x}}(8)$. The last column is joint minus single in percentage points with the original 72-comparison-adjusted interval.

Domain	Recognizer	$J(8)$	$S_{\mathrm{x}}(1)$	Image 8	$J_{\mathrm{x}}(8)$	Gain [adjusted CI]
Float32	ArcFace	0.565	0.362	0.411	0.555	19.3 $[9.4, 29.2]$
Float32	AdaFace	0.615	0.445	0.510	0.581	13.5 $[4.7, 22.4]$
Float32	MagFace	0.484	0.339	0.396	0.521	18.2 $[10.2, 26.3]$
8-bit PNG	ArcFace	0.578	0.380	0.445	0.562	18.2 $[9.6, 26.8]$
8-bit PNG	AdaFace	0.607	0.469	0.529	0.602	13.3 $[4.2, 22.4]$
8-bit PNG	MagFace	0.513	0.372	0.432	0.510	13.8 $[5.5, 22.4]$
JPEG95	ArcFace	0.036	0.161	0.201	0.242	8.1 $[1.6, 14.8]$
JPEG95	AdaFace	0.036	0.188	0.232	0.281	9.4 $[2.9, 16.4]$
JPEG95	MagFace	0.021	0.174	0.198	0.237	6.2 $[-0.5, 13.8]$

Table 16: NTU export endpoints. Frozen $J(8)$ is shown alongside domain-selected $S_{\mathrm{x}}(1)$, image averaging and $J_{\mathrm{x}}(8)$. The last column is joint minus single in percentage points with the original 72-comparison-adjusted interval.

Domain	Recognizer	$J(8)$	$S_{\mathrm{x}}(1)$	Image 8	$J_{\mathrm{x}}(8)$	Gain [adjusted CI]
Float32	ArcFace	0.289	0.188	0.266	0.266	$7.8\ [-7.0, 24.2]$
Float32	AdaFace	0.312	0.141	0.195	0.320	$18.0\ [3.1, 33.6]$
Float32	MagFace	0.188	0.148	0.148	0.180	$3.1\ [-8.6, 14.8]$
8-bit PNG	ArcFace	0.289	0.242	0.234	0.281	$3.9\ [-10.9, 18.8]$
8-bit PNG	AdaFace	0.320	0.164	0.227	0.266	$10.2\ [-7.0, 25.0]$
8-bit PNG	MagFace	0.195	0.102	0.141	0.188	$8.6\ [-4.7, 21.1]$
JPEG95	ArcFace	0.102	0.078	0.094	0.070	$-0.8\ [-10.9, 9.4]$
JPEG95	AdaFace	0.055	0.062	0.078	0.078	$1.6\ [-9.4, 12.5]$
JPEG95	MagFace	0.078	0.062	0.047	0.039	$-2.3\ [-9.4, 4.7]$

Table 17: Within-identity differences of adapted eight-versus-one gains, in percentage points with original 12-comparison-adjusted intervals. Comparisons include both export and domain-specific attack selection.

Cohort	Recognizer	PNG minus float32	JPEG95 minus PNG
FaceScrub	ArcFace	$-1.0\ [-8.6, 7.0]$	$-10.2\ [-19.3, -1.0]$
FaceScrub	AdaFace	$-0.3\ [-8.3, 7.8]$	$-3.9\ [-13.8, 6.2]$
FaceScrub	MagFace	$-4.4\ [-12.5, 3.9]$	$-7.6\ [-17.2, 2.1]$
NTU	ArcFace	$-3.9\ [-20.3, 11.7]$	$-4.7\ [-17.2, 8.6]$
NTU	AdaFace	$-7.8\ [-26.6, 7.8]$	$-8.6\ [-22.7, 6.2]$
NTU	MagFace	$5.5\ [-8.6, 19.5]$	$-10.9\ [-21.1, -2.3]$

D.5 Pre-protection translation

Horizontal and vertical shifts are sampled uniformly within amplitudes 0, 1, 2 and 4 pixels on the 64-pixel source grid, with reflection padding. The first member is unshifted and defines the reconstruction target. Blind registration uses phase correlation between each protected member's channel mean and that of the first. A response below .05, a displacement above eight pixels in either direction, or a nonfinite estimate falls back to the unregistered member. No observation is discarded. The oracle instead uses the simulated inverse translation. Table 18 retains all four amplitudes. This controlled source translation does not model natural head rotation or nonrigid facial motion.

Table 18: Pre-ISP translations on the 64-pixel grid. Triples are ArcFace/AdaFace/MagFace Rank-1. Eight-input results use no registration, blind phase-correlation registration, or oracle registration using the known simulated shifts.

Cohort	Shift	$R(1)$	Unregistered 8	Registered 8	Oracle 8
FaceScrub	0	0.359/0.453/0.333	0.531/0.576/0.479	0.526/0.586/0.471	0.531/0.576/0.479
FaceScrub	1	0.359/0.453/0.333	0.294/0.372/0.198	0.534/0.596/0.448	0.549/0.609/0.474
FaceScrub	2	0.359/0.453/0.333	0.130/0.169/0.104	0.521/0.591/0.482	0.544/0.625/0.487
FaceScrub	4	0.359/0.453/0.333	0.055/0.068/0.031	0.518/0.578/0.458	0.539/0.591/0.482
NTU	0	0.172/0.125/0.086	0.297/0.312/0.172	0.258/0.320/0.180	0.297/0.312/0.172
NTU	1	0.172/0.125/0.086	0.227/0.242/0.117	0.234/0.273/0.148	0.266/0.289/0.164
NTU	2	0.172/0.125/0.086	0.188/0.188/0.094	0.219/0.242/0.172	0.289/0.289/0.172
NTU	4	0.172/0.125/0.086	0.102/0.117/0.039	0.234/0.219/0.125	0.312/0.305/0.180

E Restoration Controls and Open-Set Evaluation

E.1 Alternative attacks and output heads

Burstormer C7 uses the official color-denoising architecture with width 48 and six input channels (RGB plus three zeros). It encodes 72,000 observations and decodes 9,000 targets, versus 19,200

targets for J. The single-input test repeats one observation in eight positions. Every query's correctness indicator agrees between one and eight inputs, giving zero empirical paired variance; wrong predicted identities can differ.

CodeFormer processes aligned 512-pixel crops, without background restoration. Fidelity is selected from 0, .5 and 1 separately for each input route on development data. DiffBIR v2.1 uses the face SwinIR/Stable Diffusion 2.1 pipeline, 512-pixel input, ten EDM DPM++ third-order SDE steps, CFG six and control scale one. Prompts and the noise-start procedure are fixed, with one draw per input, no retries, no output selection and no identity guidance. Both use bicubic resizing and the specified uint8 interfaces. Pretraining overlap is unknown. Table 19 reports their effect on $J(8)$. These are pretrained postprocessors, not generative inverses fitted to CamPro.

Table 19: CamPro restoration controls. CodeFormer and DiffBIR are applied to $J(8)$; Burstormer is fitted to the mechanism. Every observed Burstormer one/eight correctness indicator agrees, although incorrect predicted labels can differ.

Cohort	Recognizer	$J(8)$	Burst 1	Burst 8	CodeFormer	DiffBIR
FaceScrub	ArcFace	0.565	0.190	0.190	0.391	0.391
FaceScrub	AdaFace	0.615	0.255	0.255	0.440	0.417
FaceScrub	MagFace	0.484	0.159	0.159	0.417	0.383
NTU	ArcFace	0.289	0.148	0.148	0.258	0.227
NTU	AdaFace	0.312	0.141	0.141	0.242	0.266
NTU	MagFace	0.188	0.055	0.055	0.141	0.125

The Set Transformer replaces the feature moments with four-head self-attention, feed-forward width twice the feature width and a learned pooling query at each spatial location and scale. It has 3,273,955 parameters, zero dropout and no positional encoding, sharing the encoder, decoder, loss and fitting budget with the moment model. Table 20 reports both cardinalities and the adjusted intervals.

Table 20: Set Transformer endpoints, with ArcFace/AdaFace/MagFace triples. The last three columns are the original 12-comparison-adjusted intervals in percentage points for eight-minus-one gains. One model is used at both counts.

Cohort	$R(1)$	$R(8)$	ArcFace CI	AdaFace CI	MagFace CI
FaceScrub	0.344/0.424/0.341	0.510/0.576/0.492	[10.7, 22.9]	[8.1, 22.4]	[8.9, 21.4]
NTU	0.188/0.148/0.109	0.305/0.250/0.188	[3.1, 21.1]	[0.8, 18.8]	[0.8, 16.4]

The paired bounded/direct heads C5 and C6 share the parameter count and training stream. On the 64-pixel target grid, eight-input PSNR changes from 18.72 to 19.66 dB on FaceScrub and 14.12 to 16.41 dB on NTU. All six paired Rank-1 intervals include zero. These PSNRs use the head experiment's target grid and should not be substituted for metrics computed against the primary 112-pixel target.

E.2 Calibration and clean targets

Each 3,072-identity gallery contains the cohort's clean enrollment images and LFW distractors. Separate groups of 1,000 LFW unknown identities are used for calibration and evaluation. Exact FaceScrub names from all roles are excluded from these LFW pools; unrecorded aliases and recognition-model pretraining overlap remain unknown. LFW supplies unknowns for both cohorts, so calibration matches the reconstruction path, not the NTU acquisition domain.

For each recognizer and attack, the unknown score is the largest cosine similarity to the gallery. At nominal FPIR $\alpha = .01$ or .05, the threshold is the next representable value above the higher empirical $1 - \alpha$ quantile. It is fixed before target and unknown evaluation. TPIR requires both correct top rank and threshold acceptance. Table 21 uses reconstructed-domain calibration for $S_{\mathrm{d}}(1)$, $J(1)$ and $J(8)$. Table 22 evaluates genuinely clean targets with the frozen clean-domain thresholds. The clean comparison shows that the protocol already gives low NTU acceptance before protection; it does not separate gallery difficulty, domain shift and recognition quality.

Table 21: CamPro open-set results in the 3,072-identity galleries. Thresholds use reconstructed LFW unknowns, separately for each attack. Columns pair realized FPIR and TPIR at nominal 1% and 5%. Zero TPIR denotes no correctly identified and accepted target; it does not bound unseen-identity TPIR.

Cohort	Attack	Recognizer	Rank-1	FPIR 1	TPIR 1	FPIR 5	TPIR 5
FaceScrub	$S_{\mathrm{d}}(1)$	ArcFace	0.135	0.011	0.000	0.055	0.000
FaceScrub	$S_{\mathrm{d}}(1)$	AdaFace	0.232	0.006	0.000	0.060	0.016
FaceScrub	$S_{\mathrm{d}}(1)$	MagFace	0.128	0.007	0.000	0.049	0.005
FaceScrub	$J(1)$	ArcFace	0.138	0.008	0.000	0.046	0.000
FaceScrub	$J(1)$	AdaFace	0.232	0.012	0.000	0.061	0.003
FaceScrub	$J(1)$	MagFace	0.117	0.012	0.000	0.043	0.003
FaceScrub	$J(8)$	ArcFace	0.276	0.013	0.003	0.056	0.008
FaceScrub	$J(8)$	AdaFace	0.432	0.016	0.016	0.054	0.039
FaceScrub	$J(8)$	MagFace	0.234	0.008	0.005	0.039	0.010
NTU	$S_{\mathrm{d}}(1)$	ArcFace	0.016	0.012	0.000	0.060	0.000
NTU	$S_{\mathrm{d}}(1)$	AdaFace	0.016	0.004	0.000	0.060	0.000
NTU	$S_{\mathrm{d}}(1)$	MagFace	0.016	0.008	0.000	0.052	0.000
NTU	$J(1)$	ArcFace	0.016	0.010	0.000	0.041	0.000
NTU	$J(1)$	AdaFace	0.023	0.010	0.000	0.055	0.000
NTU	$J(1)$	MagFace	0.016	0.013	0.000	0.042	0.000
NTU	$J(8)$	ArcFace	0.086	0.013	0.000	0.058	0.000
NTU	$J(8)$	AdaFace	0.125	0.016	0.000	0.048	0.000
NTU	$J(8)$	MagFace	0.047	0.008	0.000	0.039	0.000

Table 22: Genuinely clean targets in the same galleries, using frozen clean-domain thresholds at nominal FPIR 1%. Parentheses give accepted-target and false-positive counts; these are not clean thresholds applied to reconstructions.

Cohort	Recognizer	Rank-1	TPIR [count]	FPIR [count]
FaceScrub	ArcFace	0.984	0.628 (241/384)	0.005 (5/1000)
FaceScrub	AdaFace	0.987	0.701 (269/384)	0.008 (8/1000)
FaceScrub	MagFace	0.982	0.578 (222/384)	0.010 (10/1000)
NTU	ArcFace	0.828	0.195 (25/128)	0.005 (5/1000)
NTU	AdaFace	0.852	0.219 (28/128)	0.008 (8/1000)
NTU	MagFace	0.664	0.062 (8/128)	0.010 (10/1000)

F Reproduction and Numerical Scope

The original synthetic inverse models and recomputation materials are in the anonymous collection. `campro-followup-materials.json` indexes retained predictions and human inverse weights, including A_0/B_8. These originals are not publicly deposited; retraining requires permitted data and fitting recipes.

Aggregate statistics. The supplement provides success totals, paired-difference histograms, COCO precision–recall arrays, saved bootstrap summaries and standard-library scripts for Tables 3, 7 and 5, associated AP/AR reductions and percentile intervals. Identity histograms determine each contrast's marginal empirical bootstrap distribution. Model inference, detection matching and regeneration of image-level resamples require upstream records outside this supplement.

Calibration numerics. Calibration maxima differ by at most 2.15×10^{-6} from the saved values; four comparisons exceed the 2×10^{-6} tolerance for an unresolved reason. Original thresholds and all target and unknown counts remain unchanged across this observed discrepancy range.

Scope of the numerical correction. The excluded optical experiments used unstable float32 factorial evaluation of high-order Zernike polynomials; their results do not validate the corrected model. CamPro does not use that basis.