Blurring Does Not Anonymise: Probing a Frozen rPPG Encoder for Identity Under Input Degradation

Tanmay Hinge
Independent researcher
hingetanmay@gmail.com

Abstract—Remote photoplethysmography (rPPG) estimates heart rate from face video and is marketed as privacy-preserving on the grounds that only a number leaves the device. That claim describes the output. It says nothing about the intermediate representation, which is what gets cached, batched, and shipped off-device. We freeze released PhysNet checkpoints, take a 1536-dimensional embedding from three internal layers, and fit linear probes for heart rate and for subject identity on the same embedding while degrading the input along an ordered ladder that averages the skin region into progressively coarser blocks. Identity is recovered at 100 % top-1 out of 120 subjects (chance 0.83 %) at every rung from a full crop down to a 4×4 block average, with verification AUC never below 0.992. It falls only when the outline of the skin region is deleted as well, to 12.5 % and AUC 0.726. The silhouette, not the texture, carries the identity. Two controls constrain what this licenses. A raw-pixel probe with no network matches the embedding down to a 4×4 block average and overtakes it below that, reaching 98.3 % where the embedding reaches 12.5 %, so the leak is a property of the input rather than of the heart-rate objective. A randomly initialised encoder of the same architecture also beats the trained one there. A pre-registered operating point requiring heart-rate error within 2 bpm of the full crop and verification AUC at or below 0.55 is empty: every condition meets the heart-rate tolerance and none approaches the identity ceiling. The heart-rate arm is null, which we report rather than suppress, and which bounds the conclusion to re-identification. The identity task links clips of one recording rather than a subject across sessions, so every figure here is an upper bound on re-identification, and we treat it as one.

Index Terms—remote photoplethysmography, biometric privacy, re-identification, linear probing, representation leakage, face de-identification

I. Introduction

A camera can measure a pulse. Each heartbeat pushes blood into the capillaries under the skin, blood absorbs green light more strongly than the surrounding tissue, and the face therefore darkens and brightens very slightly in time with the beat. The effect is far below what an eye can see and well within what a sensor can record. Averaging the colour of the skin pixels in each frame yields a waveform at the heart rate, which is the whole of the classical method [1], [2].

Contemporary systems replace that average with a convolutional network over a cropped face clip [3]. The application pitch is consistent across fitness, driver monitoring and telehealth: no contact, no wearable, and only a heart rate comes out.

That is a claim about the output. The network's intermediate activations are a separate object, and they are the object that gets cached for debugging, batched for throughput, or sent to a server when inference is not local. What those activations carry, beyond the pulse they were trained to produce, is not reported in the applied literature.

We measure it. A released rPPG checkpoint is frozen, its internal activations are pooled into a fixed-length embedding, and two linear probes are fitted on that same embedding: one for heart rate, one for subject identity. The input is then degraded along an ordered ladder that removes spatial detail from the skin region while preserving, exactly, the spatial mean that the classical methods depend on. Both quantities are read off one representation against one axis, which is what makes them comparable at all. Prior work measures heart-rate accuracy and re-identifiability on different models, different data and different scales, so the question of how much privacy buys how much signal has had no common axis.

The result is not the tidy trade-off curve the design anticipated. Our contributions:

1) Block-averaging the skin region does not reduce re-identifiability. Identity top-1 holds at 100 % out of 120 at every rung from the full 72×72 crop down to 4×4 blocks, and the result replicates on a second, independently trained checkpoint.
2) The identity signal is carried by the shape of the skin mask rather than by the pixels inside it. Replacing the masked region's average colour with the same colour painted across the whole frame, which leaves three numbers per frame, drops identity from 75.8 % to 12.5 %.
3) The encoder discards identity that its own input retains. On that condition a raw-pixel probe reaches 98.3 % and an untrained encoder 44.2 %, both above the trained encoder's 12.5 %.
4) A pre-registered privacy–utility operating point does not exist anywhere in the sweep, and the reason is one-sided: the heart-rate criterion is met everywhere and the identity criterion is met nowhere.

We also report a null result in full. The heart-rate probe

does not separate from the baseline of ignoring the video and guessing the dataset mean, which limits the trade-off reading of the empty operating point and is addressed directly in Section VI.

II. Related Work

A. Remote photoplethysmography

Verkruysse et al. [1] established that ambient-light video carries a plethysmographic signal. Chrominance-based CHROM [4] and the plane-orthogonal-to-skin method POS [5] derive the pulse analytically from the spatially averaged skin colour, with no learned parameters. Learned estimators such as PhysNet [3] predict the waveform directly from a face clip. We use both families: PhysNet as the encoder under test, and POS and CHROM as floors that say what a method with no training achieves on identical data.

B. Obfuscation does not anonymise

That blurring and pixelation fail as privacy controls is established outside this domain. Newton et al. [6] showed that naive de-identification leaves faces linkable. McPherson et al. [7] recovered identities from mosaiced, blurred and encrypted images using ordinary networks, and Oh et al. [8] recognised people whose faces were entirely obscured. This literature has not been connected to rPPG, where blurring is proposed as a defence.

C. Privacy-preserving rPPG

Closest to this work, Gupta and Etemad [9] perturb the input by extracting face regions with less identity content, shuffling pixels under a per-video key held constant across frames, and blurring, and report face-recognition accuracy falling substantially while heart-rate error moves little.

Two differences matter. The first is the attacker. They embed the perturbed image with ArcFace [10] and classify it, which is the correct test against someone who intercepts the video and no test at all against someone who obtains the network's own embedding. A defence can be sound against the first attacker and untested against the second. The second is the direction of their own sweep: it selects a 3×3 kernel because that size is optimal for suppressing identification, which implies that larger kernels recover it. Their method composes three operations, so what we report tests components rather than their pipeline. The components are what a practitioner reaches for.

D. Probing

Fitting a linear model to frozen activations to ask what a representation contains [11] is standard, and so is the accompanying caution: an accuracy is uninterpretable without the floors it must clear. We adopt the control-task construction of Hewitt and Liang [12] and report selectivity, the gap between a probe's score and its score on randomised labels, alongside every result.

III. Method

A. Threat model

The adversary holds the encoder's intermediate representation for a set of clips and nothing else: not the video, not the pulse waveform, not the model's gradients. This is the position of anyone with access to a debugging cache, an inference log, a batch in transit to a server, or a stored feature database, and it is the position that the privacy-preserving framing of rPPG does not address, because that framing is a statement about the output.

The adversary's capability is deliberately weak. They fit a linear model, or in the verification case fit nothing at all and compare cosine similarities. No fine-tuning, no inversion, no auxiliary network. What such an adversary recovers is therefore a lower bound on what the representation exposes, while what it recovers about identity is an upper bound on re-identification for the separate reason given in Section VIII.

This differs from the attacker assumed by perturbation defences, who receives the degraded image and runs a face recogniser on it. The two are not interchangeable, and Section II-C returns to the point.

B. Data

SCAMPS [13] provides synthetic avatar videos rendered with a known pulse driving the skin colour, so ground-truth heart rate is exact rather than measured with a contact sensor. Every face in this paper is a rendered avatar and no human subject appears. We use 120 videos of 600 frames at 30 fps, cut into non-overlapping clips of 128 frames, giving 480 clips. Overlapping clips would leak frames across the identity probe's split, so the stride equals the clip length. One video was excluded: the face detector reported success on a patch of wall containing no skin, and a mask-fraction check caught it. Section IX shows the case.

Frames are cropped to the face and resized to 72×72, which is PhysNet's training input size and not a free parameter. Per-frame skin masks supplied with the dataset are reduced to one static mask per video. Everything outside it is zeroed in every condition except the unmasked control.

C. Encoders

We use released PhysNet checkpoints from the rPPG-Toolbox [14], pinned at commit b7500b848f with per-file hashes recorded (the primary checkpoint is cf295ff0b7a43ee6), loaded strictly so that a key mismatch cannot leave part of a nominally pretrained encoder randomly initialised. The encoder is set to evaluation mode and its gradients disabled. Evaluation mode matters independently of gradients: PhysNet's batch normalisation running statistics would otherwise be updated by the very forward passes done to read activations, making the encoder a function of how many clips it had already seen.

The primary encoder is trained on PURE [15], a dataset of real subjects, so it provably never saw the evaluation data. The replication arm is trained on UBFC-rPPG [16], also real subjects, with no shared training data. A third checkpoint trained on SCAMPS itself is reported only in the appendix, because most of the videos scored here fall in its training split.

D. Embedding

Forward hooks on three internal convolutional blocks capture activations that are pooled over time by mean and standard deviation on a spatial grid, giving 1536 dimensions per clip. The standard-deviation half is not optional. The pulse is the temporal variation of these activations, so mean-only pooling would act as a low-pass filter that keeps static appearance and discards pulsatility, biasing every result in exactly the direction under test, invisibly.

E. The degradation ladder

The ladder removes spatial detail from the skin region in ordered steps: the full crop, then mask-weighted averages over progressively coarser cells, down to a single average colour per frame and channel filling the mask. Nothing about the pulse is deliberately removed on the way down. Averaging pixels together does not weaken a signal that every pixel shares.

This is enforced rather than assumed. Every ladder condition preserves the mask-weighted spatial mean per frame and channel to within 3.4×10^{-3} on a 255-level scale, against a pulse amplitude of several units on that same scale. The property is what makes the ladder a spatial-detail control rather than a signal-attenuation control, and Section IV turns it into an end-to-end check on the implementation.

Four conditions sit off the ladder and answer different questions: green channel only; pixel positions shuffled within the mask under one permutation held constant across a clip's frames, so each pixel keeps its own history while the layout is destroyed; the untouched crop with hair and background retained, which measures how much identity lives outside the skin region; and the single average colour broadcast over the entire frame with no outline, which is the condition that turned out to matter most. Fig. 1 shows all ten.

F. Probes

Four probes read the same embedding. Heart rate is estimated by ridge regression ($\alpha = 1.0$) on standardised embeddings, reported as mean absolute error in bpm. The encoder's own rPPG decoder output provides a probe-free reference. Identity is measured two ways. Closed-set identification fits multinomial logistic regression (at most 1000 iterations) over 120 classes on 360 clips and evaluates on the 120 held out, one per video and always that video's temporally last. Probe capacity is kept low deliberately: the argument for probing is that a linear model is too weak to manufacture structure the representation does not contain. Verification requires no fitting: clip embeddings are L2-normalised, cosine similarity is computed over all 114,960 pairs, and same-subject and different-subject pairs are separated by ROC AUC. Verification is the better-powered metric and the one that matches the threat, since it re-identifies subjects the probe was never fitted on.

Heart-rate frequency estimation uses a zero-padded periodogram over 0.75–2.5 Hz. Padding is not cosmetic: an unpadded transform on a clip of this length resolves only to 14.06 bpm, so every returned value lands on a coarse grid, which Section IX shows. A study about a few bpm cannot be measured that way.

IV. Controls

An accuracy without a floor is not a measurement. Every result carries four.

Untrained encoder. The identical architecture with random weights, over three seeds, since one random initialisation is itself a draw. This says what any arbitrary projection of the input would score.

Raw pixels. The same probe fitted directly on the degraded frames, with no network in the loop, combining a coarse spatial pooling of the temporal-mean image with the mask-mean colour trace. This says whether the encoder was needed. It is computed from the same degraded frames the encoder saw, never the clean ones.

Permuted labels. Targets drawn deterministically per clip from the real target's own marginal, and fixed once rather than resampled, so that the probe faces a task with definitionally nothing to learn. Selectivity, the difference between the real and permuted scores, is reported rather than a bare accuracy.

POS and CHROM. Analytic methods with no learned parameters, which say whether a decade of representation learning bought anything on this data.

The last of these also yields a validity check on the whole pipeline. POS and CHROM reduce each frame to its spatial mean before doing anything else, and Section III-E guarantees that every mean-preserving condition leaves that mean untouched. Both methods must therefore return identical heart rates at the full crop, at every block level, at the single trace, and under shuffling. They do, to the digit (130.078125 bpm). If the masking, the block partition, the alignment between mask and frame after cropping, or the clip assembly were wrong anywhere, that equality would break. One assertion covers most of the implementation.

V. Results

A. Identity survives the ladder intact

Chance on this task is 0.83 %. Identity is 100 % at the full crop and 100 % at a 4×4 block average: sixteen coloured squares per frame, one of 120 subjects, and the probe makes no error. Verification AUC over the same range falls only from 0.9998 to 0.992. The entire ordered

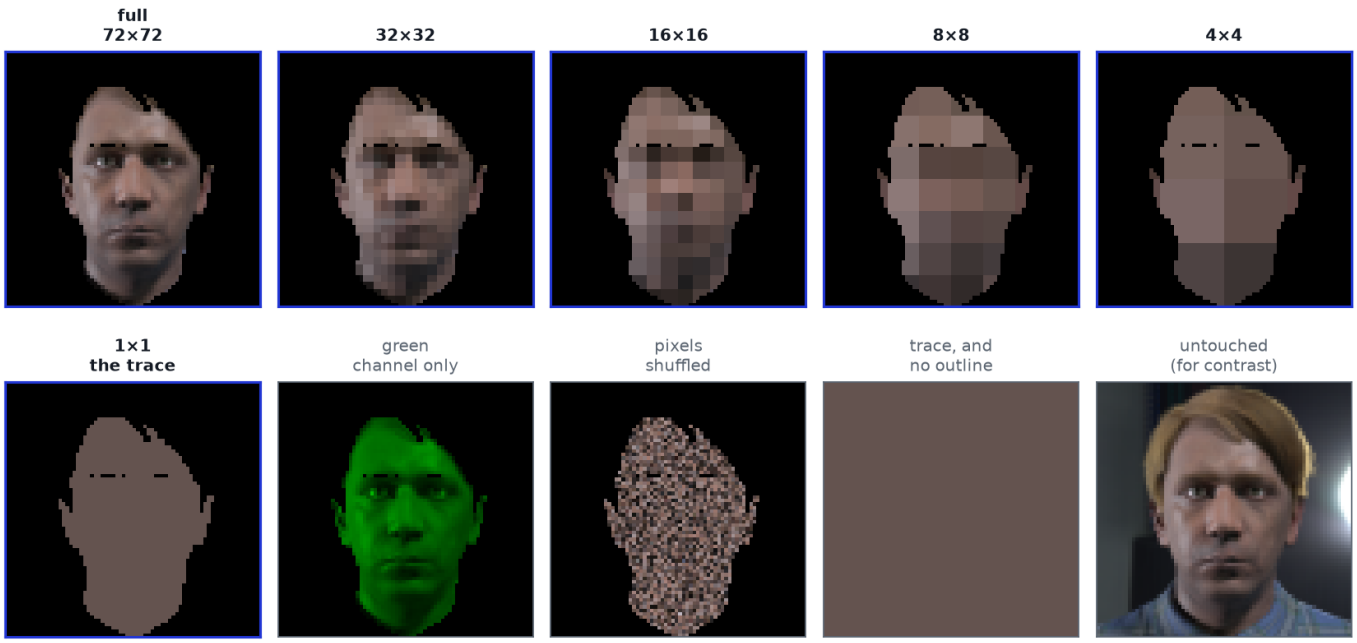


Fig. 1. The ladder (top) and the four off-ladder conditions (bottom). A 4×4 block average is sixteen coloured squares and is still recognisably a face-shaped object. The last panel of the bottom row removes that shape.

ladder costs essentially nothing in re-identifiability (Fig. 2, top).

Cosine nearest-neighbour, which fits nothing at all, agrees down the ladder: 100 % at the full crop and 97.5 % at 4×4 blocks. The ladder result is therefore a property of the embedding rather than of the classifier's capacity.

Shuffling pixel positions within the mask leaves identity at 99.2 %, which already indicates that the probe is not relying on facial layout. The permuted-label control scores 0.8 %, chance to the reported precision, so the perfect score is not a high-capacity probe memorising its way to any answer.

Selectivity, the gap to that control, is 0.99 at the full crop.

The two remaining off-ladder conditions answer the questions they were built for, and both answers are negative for privacy. Retaining hair and background instead of masking to skin leaves identity at 100 % and AUC 1.000, so restricting the model's input to the skin region removes nothing. Keeping only the green channel also leaves identity at 100 %, while POS degrades from 15.8 to 50.4 bpm on those same clips, because a chrominance method needs all three channels. Discarding colour information costs the pulse and not the person.

Because identification holds out exactly one clip per video, these scores are counts of successes over 120 independent trials, one per video, and admit exact Wilson intervals: 97–100 % at the full crop and 97–100 % at 4×4 blocks. The comparison against chance is not close enough for an interval to alter it.

B. The silhouette is the leak

At the bottom rung the entire skin region is one average colour per frame, and identity is still 75.8 % (91 of 120, Wilson 67.4–82.6 %), about 91 times chance. One colour should not support that.

Here the two probes part company. Nearest neighbour recovers only 38.3 % on this condition against the fitted classifier's 75.8 %, so a substantial part of the single-trace result is capacity the classifier supplies rather than structure sitting in the embedding. The finding below survives that distinction, because it turns on a comparison between two conditions measured with the same probe.

It is not one colour. It is one colour in the shape of that subject's face, with holes where the eyes and eyebrows are, because everything outside the skin region is zeroed and the skin region is a silhouette. The condition removed all texture and left a stencil.

Broadcasting that same average colour across the whole frame, which leaves three numbers per frame and no outline, drops identity to 12.5 % (15 of 120, Wilson 7.7–19.6 %) and verification AUC to 0.726 (Fig. 3). The two conditions do not overlap: 67.4–82.6 % against 7.7–19.6 %. That is the finding. Blurring does not anonymise a face. Blurring together with deleting the shape of the face gets most of the way there, and even then 12.5 % out of 120 subjects is about 15 times chance.

C. The encoder discards what its input keeps

The raw-pixel floor matches the embedding at 100 % down to 4×4 blocks. Below that it stops matching, and in the unexpected direction. At the single trace the

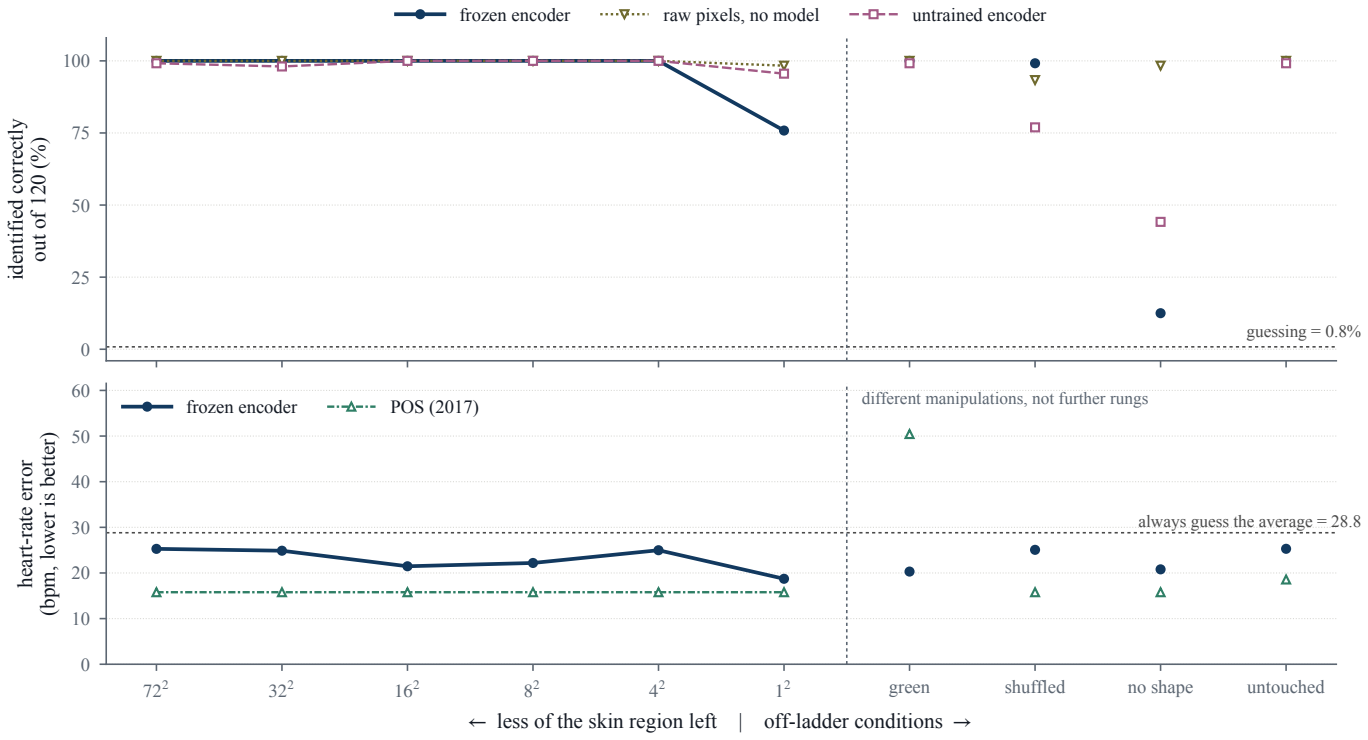


Fig. 2. Both quantities read off one embedding against one axis. Identity (top) does not move until the bottom of the ladder. Heart-rate error (bottom) never departs far from the error of ignoring the video and guessing the dataset mean. Ladder rungs are joined because they are ordered. The off-ladder conditions are drawn as bare markers because they are not a sequence.

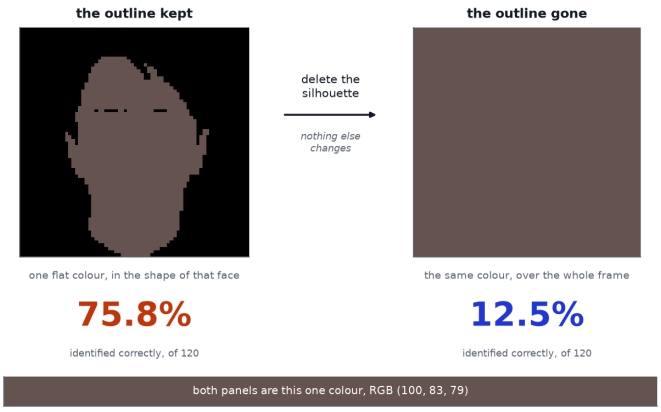


Fig. 3. The same colour, twice. Left, the mask outline retained: 75.8 % identified. Right, the identical colour filling the frame: 12.5 %. Every pixel on the left is either that colour or black, so the identifying information is in where the boundary falls.

embedding is already behind (75.8 % against 98.3 %). With the outline deleted the embedding falls to 12.5 % while the same probe on the identical input, with no network at all, holds 98.3 %. On verification that is 0.726 against 1.000. Passing the image through the encoder destroyed identity that survived the degradation itself.

Nearest neighbour on the same embedding gives 8.3 %, so this is not a capacity effect in either direction. The untrained encoder behaves the same way: 44.2 % closed-set and 0.942 AUC, both above the trained encoder. A randomly initialised network separates these subjects better than one that spent its training learning to read pulses.

A plausible mechanism, offered as a hypothesis because nothing here isolates it: both checkpoints consume frame differences, and a spatially flat frame has almost none. The constant skin tone is normalised away before the first convolution, while raw pixels retain the absolute colour that on rendered avatars is close to a fingerprint.

D. Replication on a second checkpoint

Table I repeats the identity measurements on the UBFC-trained checkpoint. The two encoders share no training data and no fitting. Both hold at a perfect score through the coarse block averages, both slip at the single trace, and both fall at the same place, the moment the silhouette is removed. They disagree on the depth of the fall (12.5 % against 24.2 %) but not on its location or its cause. Two models trained on different people degrade the same way, because what is being destroyed is in the input rather than in either of them.

E. The heart-rate arm is null

The lower panel of Fig. 2 is nearly flat, which could be described as heart rate surviving the degradation. That would be a fair description of the shape and a misleading

TABLE I
Identity, top-1 out of 120, on two independently trained encoders. Chance is 0.83 %.

condition	PURE	UBFC-rPPG
72^2	100.0	100.0
8^2	100.0	100.0
4^2	100.0	96.7
1^2	75.8	62.5
no shape	12.5	24.2

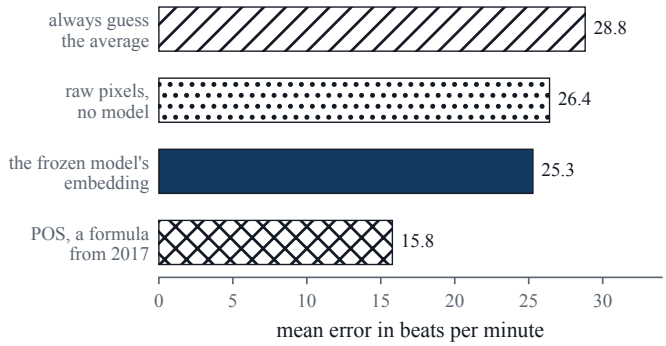


Fig. 4. Heart-rate error at the full crop against the three floors it must clear. The frozen encoder's embedding is better than nothing and worse than an analytic method from 2017.

description of the result, because the line is flat at a level that is not an achievement.

Ignoring the video entirely and always answering with the dataset's mean heart rate is wrong by 28.8 bpm. The probe on the frozen embedding reaches 25.3 bpm. POS reaches 15.8 bpm on the same clips and CHROM 21.5 bpm (Fig. 4). Bootstrapped over videos, the probe's interval runs from 22.1 to 28.7 bpm, and the upper end of that interval is the value obtained for free by ignoring the video. The model's advantage over guessing is not something this study can distinguish from zero.

The encoder's own decoder, read without any probe, tells the same story and adds one detail. It reaches 19.8 bpm at the full crop, but its median absolute error is 1.5 bpm and 57 % of videos land within 5 bpm. The mean is therefore set by a minority of videos that fail badly rather than by uniform jitter, which is what the harmonic diagnostics in Section VIII describe.

The likely reason is domain shift: the checkpoints were trained on real subjects and evaluated here on synthetic renders. Faces built by a graphics engine are not the same distribution as faces lit by a room, and the pulse is the component that would suffer first, being a sub-percent colour fluctuation that a renderer approximates. Identity, which lives in gross appearance, transfers without difficulty.

VI. The Operating Point That Does Not Exist

A camera product that reads pulses and wants to be honest about privacy needs a rung of this ladder it can actually run at: one that keeps heart-rate accuracy close to the full crop and pushes re-identification towards chance. Written down in advance, that was heart-rate error within 2 bpm of the full crop and verification AUC at or below 0.55.

Every condition in the sweep, on the ladder and off it, meets the heart-rate tolerance. The pulse never got appreciably worse, because there was not much of it there to lose. The question therefore reduces to the identity criterion, and that one is not close. The best any ladder rung reaches is 0.898. Leaving the ladder entirely and deleting the outline down to three numbers per frame reaches 0.726. The target was 0.55 (Fig. 5). The search records its answer as empty.

TABLE II
All conditions, PURE PhysNet DiffNormalized. Identity is top-1 over held-out clips, by the fitted probe and by cosine nearest neighbour, against the raw-pixel floor; chance is 0.83 %. Heart-rate columns are mean absolute error in bpm.

	identity, top-1 (%)			heart rate (bpm)	
condition	probe	1-NN	raw px	probe	POS
72^2	100.0	100.0	100.0	25.3	15.8
32^2	100.0	100.0	100.0	24.9	15.8
16^2	100.0	99.2	100.0	21.5	15.8
8^2	100.0	98.3	100.0	22.2	15.8
4^2	100.0	97.5	100.0	25.0	15.8
1^2	75.8	38.3	98.3	18.7	15.8
green	100.0	100.0	100.0	20.3	50.4
shuffled	99.2	96.7	93.3	25.1	15.8
no shape	12.5	8.3	98.3	20.8	15.8
untouched	100.0	100.0	100.0	25.3	18.6

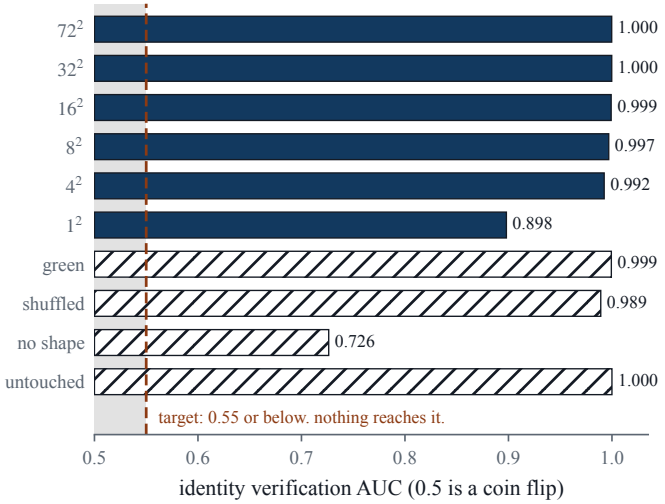


Fig. 5. Verification AUC per condition, measured up from a coin flip. The shaded sliver is the target region. The shortest bar in the study is several times longer than the target is wide, and every condition already satisfies the heart-rate criterion, so this is the only axis on which the search fails.

A. Two readings, and what separates them

An empty target region is consistent with two different worlds. In the first, pulse and identity are genuinely entangled in this representation and no degradation of this kind separates them. In the second, this encoder never transferred to these faces, the heart-rate axis was dead on arrival, and the observation that the pulse does not survive degradation restates the fact that there was no pulse in the embedding to begin with. The heart-rate numbers in Section V-E are equally consistent with both.

One piece of evidence narrows it. POS, which has no parameters and no training, recovers heart rate from these same videos at 15.8 bpm, so the pulse is present in the pixels and nothing about the dataset makes it unrecoverable. What failed is the frozen encoders' transfer onto rendered faces, not the data. That rules out the strongest form of the second reading without closing it, because an encoder that cannot read the pulse at all cannot be asked to trade it away.

Settling it requires an arm on real data where these checkpoints are in domain. If the target region is still empty once the heart-rate axis is alive, the coupling claim stands. If a rung opens up, then this study measured a broken encoder and reported something about the ladder instead. That experiment has not been run. Until it has, the identity result stands on its own and the trade-off claim is provisional.

VII. Discussion

A. Blur strength is not a privacy parameter

The practical reading is narrow and firm. Across every level of block averaging tested, on two independently trained encoders, coarsening the skin region bought nothing in re-identifiability while the pulse was never the binding constraint. A system that ships blur as its privacy control, and that reports a kernel size as though it were a privacy budget, is reporting a quantity that on this evidence does not vary with the thing it is supposed to protect. That is consistent with the obfuscation literature outside this domain [7], [8] and, to our knowledge, has not previously been checked inside it.

B. Masking to the skin region adds a leak

The silhouette result cuts against a design pattern that is close to universal in rPPG pipelines. Restricting the input to a skin mask is done for signal quality, and it is usually described, when privacy is discussed at all, as a reduction in what the model sees. It is not. The mask boundary is subject-specific geometry, and at the bottom of the ladder it is the only thing left carrying identity. Zeroing outside a per-subject mask writes that subject's outline into every frame. A pipeline that both masks and blurs has removed the texture and kept the stencil.

C. Image-space evaluation does not bound representation-space leakage

Encoders are not uniformly worse than the pixels they consume. At the bottom of the ladder this encoder discards identity that its own input retains, and a randomly initialised network of the same architecture retains more of it than the trained one. The direction is surprising, but the consequence for evaluation does not depend on the direction: the two quantities are not the same measurement, so a defence validated by running a face recogniser on the perturbed image has not been validated against an adversary holding the representation. Both should be reported. A defence may well be sound against both, and the present evidence cannot say which way a given method falls, only that one result does not substitute for the other.

D. What this does not license

It does not establish that pulse and identity are inseparable in principle. The empty operating point rests on an encoder whose heart-rate axis never came alive on this data, and Section VI sets out the experiment that would separate the two readings. It does not establish a re-identification rate, for the reason given below. And it says nothing about behaviour across real skin tones, which needs a dataset that labels them.

VIII. Limitations

The identity task is easier than re-identification. The dataset gives no correspondence between avatars across videos, so "the same subject" here means a different clip of the same recording: same avatar, same lighting, same camera. A probe that separates those has done something easier than recognising someone who returned on another day. Every identity number in this paper is an upper bound rather than an estimate. Datasets that would make it a fair estimate, PURE with subjects recorded across sessions and MMPD [17] with subjects across lighting conditions, are the obvious next step, and the analysis code is written against an interface that admits them.

Verification AUC has no honest interval. Closed-set identification does, because it is a count over independent per-video trials, and those Wilson intervals are reported with the results. Verification is different: 480 clips generate 114,960 pairs of which 720 are positive, and each clip appears in hundreds of them. The analytic Hanley–McNeil interval for the lowest AUC in the study is 0.705–0.747, but it assumes one observation per subject and so understates the width; it is a lower bound on uncertainty rather than a measurement of it. A correct interval needs the per-pair similarities resampled with the video as the cluster, and the committed run recorded only the aggregate. The operating point is defined on this metric, which makes this the study's weakest statistical link.

Synthetic data supports no claim about skin tone. Skin tones here are whatever the renderer produced. Nothing

in this paper says how any of this behaves across real skin tones. That question needs a dataset with Fitzpatrick labels, and the analysis code refuses to pool heart-rate error across those groups when they are present, because reporting one average across skin tones is how a method that works for some people and not others gets hidden.

Masked input is off-distribution. Zeroing everything outside the skin region is not what these checkpoints were trained on, so the absolute heart-rate errors here are not comparable to published numbers on this dataset. The comparisons within this study, which all share that treatment, are unaffected.

Error is dominated by harmonics. At low heart rates the second harmonic falls inside the analysis band and can exceed the fundamental. Both PhysNet and POS make this error on the same videos, which is why median error and the fraction within 5 bpm are reported alongside the mean.

The subject count is a bandwidth budget, not a power calculation. The dataset's per-file download links are dead and the only working route is a single very large compressed archive that cannot be seeked into, so the sample is a prefix of that archive rather than a random draw.

IX. Reproducibility

The study is config-driven and seeded, runs on a laptop, and is covered by 121 tests with the degradation ladder and the probing logic as explicit acceptance gates against written specifications. Preprocessing is checked to be bit-identical to the reference toolbox implementation, so that a reimplementation did not quietly become a modification. The invariance property in Section IV is asserted by the test suite rather than merely observed.

Three implementation errors are worth recording because each produced plausible output. The frequency estimator returned values on a 14.06 bpm grid, which is invisible until the returned values are plotted and noticed to be suspiciously round (Fig. 6). The face detector reported success on a patch of patterned wall behind an avatar wearing a head-mounted display, and the resulting crop contained no skin at all, and nothing downstream would have failed, because the averages would simply have been zeros (Fig. 7). And holding every clip's frames in memory drove the sweep into swap, which a change to memory-mapped views fixed at a cost of one line.

X. Conclusion

Degrading a face by block-averaging its skin region does not make the subject less identifiable from a frozen rPPG encoder's internal representation, at any level of degradation tested, on two independently trained checkpoints. What carries the identity is the outline of the skin region, and removing that is what finally reduces it. Controls place the leak in the input rather than in the heart-rate objective: raw pixels match the encoder

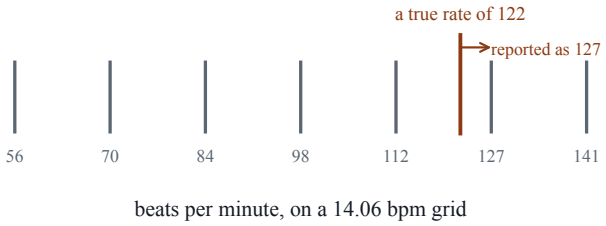


Fig. 6. Every heart rate the unpadded estimator can return for a clip of this length. A true rate between grid points has nowhere to land except one of the ticks.



Fig. 7. The detector reported a face. The red box is what it found. The right panel is what the model was then asked to measure a heart rate from. The video is excluded, with its reason recorded.

throughout and beat it once the outline is gone, as does an untrained network. The practical consequence is that blur strength is not a privacy parameter, and that a defence evaluated against an image-space face recogniser has not been evaluated against an attacker holding the representation. The pre-registered operating point is empty, though the null heart-rate arm means that result is about re-identification with confidence and about the trade-off only provisionally.

Data Availability

Results, the run manifest, per-condition scores and the operating-point record are committed to the repository at https://github.com/tanmayhinge/rppg-identity-tradeoff, together with the code that generates every number and figure in this paper from them. SCAMPS is available from its authors under the Research Use of Data Agreement.

Ethics

No human subjects were involved. Every face shown or measured is a synthetic avatar rendered by SCAMPS and corresponds to no person. The dataset is used under the Research Use of Data Agreement v1.0, which permits the small excerpts reproduced in figures that explain work done with it. The model checkpoints and the analytic baselines are used non-commercially for research. This work characterises a privacy weakness in a class of sensing system and proposes no method for exploiting it.

TABLE III
Identity top-1 (%) and heart-rate MAE (bpm) for all conditions and all three checkpoints. The SCAMPS row group is contaminated; see text.

	identity top-1 (%)			heart-rate MAE (bpm)		
condition	PURE	UBFC	SCAMPS	PURE	UBFC	SCAMPS
72^2	100.0	100.0	99.2	25.3	22.2	5.0
32^2	100.0	100.0	98.3	24.9	21.0	7.0
16^2	100.0	100.0	99.2	21.5	21.9	10.6
8^2	100.0	100.0	100.0	22.2	22.6	17.3
4^2	100.0	96.7	99.2	25.0	18.8	21.4
1^2	75.8	62.5	76.7	18.7	14.9	17.3
green	100.0	100.0	99.2	20.3	22.3	9.7
shuffled	99.2	99.2	95.0	25.1	19.9	7.5
no shape	12.5	24.2	7.5	20.8	15.7	19.8
untouched	100.0	100.0	100.0	25.3	21.6	5.0

Author Contributions

T.H.: conceptualisation, methodology, software, formal analysis, investigation, data curation, writing (original draft and revision), visualisation.

Conflict of Interest

The author declares no competing interests.

Funding

This research received no external funding.

AI Use Statement

AI was used to assist with text editing, coding, and with literature research. AI was not involved in research ideation, the development of theoretical results, research methodology, and experimental design. All AI-assisted outputs were reviewed, revised where necessary, and verified by the authors.

Appendix A
All Encoders

Table III reports every condition on all three checkpoints. The SCAMPS-trained checkpoint is not comparable to the other two: it was fitted on this dataset's own training split, and 81 of the 120 videos scored here fall in that split, so it is being evaluated partly on data it was trained on. Its heart-rate figures in particular (5.03 bpm at the full crop, against 25.3 and 22.2 for the two real-data checkpoints) are consistent with memorisation and should not be read as in-domain performance. It is included because its identity behaviour at the bottom of the ladder (7.5 %) agrees with the other two, and that agreement is the point.

References

[1] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," Optics Express, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[2] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation," Optics Express, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[3] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in Proceedings of the British Machine Vision Conference (BMVC), 2019, arXiv:1905.02419.

[4] G. de Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," IEEE Transactions on Biomedical Engineering, vol. 60, no. 10, pp. 2878–2886, 2013.

[5] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, "Algorithmic principles of remote PPG," IEEE Transactions on Biomedical Engineering, vol. 64, no. 7, pp. 1479–1491, 2017.

[6] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 2, pp. 232–243, 2005.

[7] R. McPherson, R. Shokri, and V. Shmatikov, "Defeating image obfuscation with deep learning," 2016.

[8] S. J. Oh, R. Benenson, M. Fritz, and B. Schiele, "Faceless person recognition: Privacy implications in social media," in European Conference on Computer Vision (ECCV), 2016, arXiv:1607.08438.

[9] D. Gupta and A. Etemad, "Privacy-preserving remote heart rate estimation from facial videos," in IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2023, arXiv:2306.01141.

[10] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4685–4694.

[11] G. Alain and Y. Bengio, "Understanding intermediate layers using linear classifier probes," 2016.

[12] J. Hewitt and P. Liang, "Designing and interpreting probes with control tasks," in Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019, arXiv:1909.03368.

[13] D. McDuff, M. Wander, X. Liu, B. L. Hill, J. Hernandez, J. Lester, and T. Baltrusaitis, "SCAMPS: Synthetics for camera measurement of physiological signals," 2022.

[14] X. Liu, G. Narayanswamy, A. Paruchuri, X. Zhang, J. Tang, Y. Zhang, S. Sengupta, S. Patel, Y. Wang, and D. McDuff, "rPPG-Toolbox: Deep remote PPG toolbox," 2022.

[15] R. Stricker, S. Müller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 2014, pp. 1056–1062.

[16] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," Pattern Recognition Letters, vol. 124, pp. 82–90, 2019.

[17] J. Tang, K. Chen, Y. Wang, Y. Shi, S. Patel, and D. McDuff, "MMPD: Multi-domain mobile video physiology dataset," 2023.