Re-Identification in Chest Radiographs After Anatomical Masking — Supplementary Material

S1 Cohort construction

Tables S1 and S2 summarize cohort construction and the patient-level data splits.

Table S1 | Cohort filtering. Capping removed the both-view property for a small number of patients, who were then excluded.

Filtering stage	Patients	Images
ChestX-ray14, all	30,805	112,120
≥ 2 images per patient	13,302	94,617
Both PA and AP views present	7,123	—
Capped at 10 images per patient	6,936	44,044

Table S2 | Split composition. Splits are disjoint at the patient level.

Split	Patients	AP	PA	Total
Train	4,836	14,691	15,758	30,449
Validation	700	2,166	2,319	4,485
Test	1,400	4,499	4,611	9,110

Table S3 | Patient demographics across the study cohorts. Age and sex distributions are calculated per patient rather than per image. The final row shows the full ChestX-ray14 population for comparison.

Split	Patients	Images	Age, mean $\pm$ SD	Male (%)
Train	4,836	30,449	47.37 ± 16.65	56.97
Validation	700	4,485	47.13 ± 16.71	59.43
Test	1,400	9,110	46.88 ± 16.66	58.07
Cohort (all)	6,936	44,044	47.25 ± 16.66	57.44
ChestX-ray14 (all)	30,805	112,120	45.53 ± 16.68	53.98

S2 Training hyperparameters

AdamW, learning rate 3×10^{-4}, weight decay 10^{-4}; one-cycle schedule with warmup fraction 0.1; batch size 64; 30 epochs; mixed precision; input 224×224 normalised to $[-1, 1]$. Augmentation: random affine ($\pm 5^{\circ}$ rotation, $\pm 3\%$ translation, scale 0.95–1.05) and brightness/contrast jitter of ± 0.15. No horizontal flip.

Final training loss: ResNet-50, 0.0029 / 0.0018 / 0.0032 for seeds 0/1/2; DenseNet-121, 0.0036 / 0.0023 / 0.0038.

S3 Segmentation

Table S4 reports the mean area of each of the 14 structures produced by the ChestX-Det segmentation model, at the 320×320 resolution used throughout the ablation study. For the ten retained regions, areas are averaged over the cached masks of all 9,110 test images, which are the masks that defined every ablation reported in this paper. The four excluded regions were not retained in the cache, so their areas were measured by re-running the segmentation model on a random sample of 200 test images. Both hilus pulmonis masks and the aorta mask were excluded as too small for graded ablation: each covers about 1% of the image. The weasand mask was excluded as anatomically ambiguous on plain radiography. Figure S1 shows the segmented structures on one example radiograph.

Table S4 | Mean mask area of all 14 structures segmented by the ChestX-Det model, at the 320×320 resolution used for ablation. Retained regions are averaged over all 9,110 test images; excluded regions over a 200-image sample. The image is $320 \times 320 = 102{,}400$ pixels.

Status	Region	Mean area (px, 320×320)
Retained	Diaphragmatic surface	18,791
	Left lung field	16,008
	Right lung field	15,175
	Spine	9,755
	Heart	7,380
	Left scapula	5,832
	Right scapula	5,707
	Mediastinum	4,833
	Left clavicle	1,446
	Right clavicle	1,352
Excluded	Weasand	2,654
	Right hilus pulmonis	1,216
	Aorta	1,154
	Left hilus pulmonis	1,000

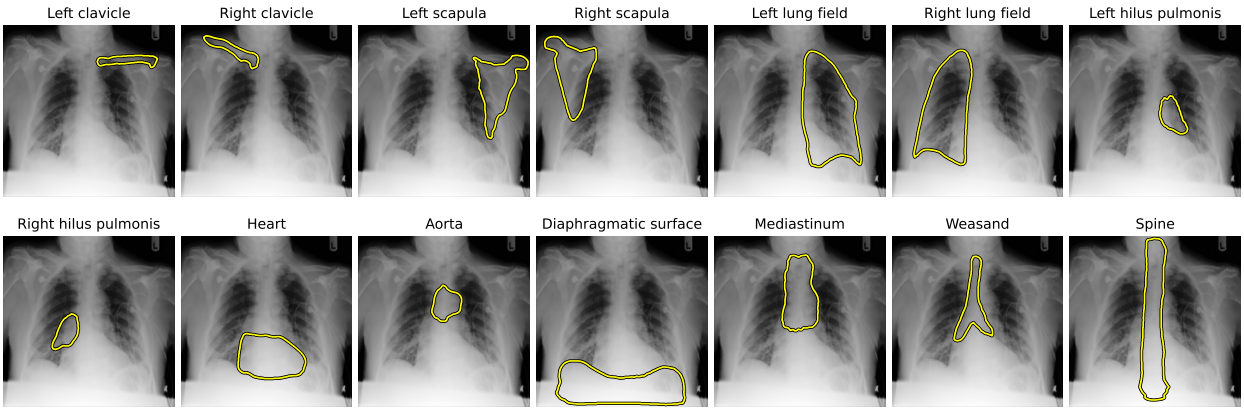


Figure S1 | All 14 segmented structures overlaid on one radiograph, shown for visual verification of mask quality.

S4 Full baseline results

Table S5 | Closed-gallery retrieval on 1,400 test identities unseen during training (ResNet-50, seed 0). Same-view conditions exclude self-matches and are restricted to patients with at least two images in that view (1,031 for PA, 957 for AP); the gallery is restricted the same way.

Condition	Gallery	p@1	top-5	top-10	top-20	mAP@R
AP $\rightarrow$ PA (primary)	4,611	0.9500	0.9764	0.9824	0.9864	0.9125
PA $\rightarrow$ AP	4,499	0.9703	0.9883	0.9918	0.9935	0.9198
PA $\rightarrow$ PA	4,242	0.9934	0.9960	0.9974	0.9974	0.9802
AP $\rightarrow$ AP	4,056	0.9731	0.9862	0.9896	0.9929	0.9076

S5 Statistical detail

Table S6 | ResNet-50, seed 0. Patient-clustered bootstrap intervals at complete regional ablation against the shape-matched control (20,000 resamples of 1,400 identities, all queries of each sampled identity included). Bonferroni intervals target a nominal family-wise error rate of 5% across ten regional comparisons. Seven of ten regions remain significant after correction. These intervals reflect sampling variability within a single trained model and are not directly comparable to the across-seed standard deviations in Table 5 of the main text.

Region	Δ	Uncorrected 95% CI	Bonferroni 99.5% CI	Sig.
Right lung field	-0.0914	$[-0.1106, -0.0722]$	$[-0.1191, -0.0636]$	yes
Left lung field	-0.0789	$[-0.0977, -0.0598]$	$[-0.1063, -0.0518]$	yes
Mediastinum	-0.0453	$[-0.0573, -0.0335]$	$[-0.0624, -0.0279]$	yes
Spine	-0.0171	$[-0.0306, -0.0037]$	$[-0.0365,\ 0.0020]$	no
Right clavicle	-0.0013	$[-0.0077,\ 0.0050]$	$[-0.0107,\ 0.0076]$	no
Left clavicle	$+0.0153$	$[\ 0.0096,\ 0.0213]$	$[\ 0.0071,\ 0.0238]$	yes
Right scapula	$+0.0184$	$[\ 0.0050,\ 0.0317]$	$[-0.0007,\ 0.0372]$	no
Left scapula	$+0.0204$	$[\ 0.0079,\ 0.0331]$	$[\ 0.0026,\ 0.0389]$	yes
Heart	$+0.0520$	$[\ 0.0405,\ 0.0635]$	$[\ 0.0354,\ 0.0693]$	yes
Diaphragmatic surface	$+0.2836$	$[\ 0.2677,\ 0.2999]$	$[\ 0.2603,\ 0.3070]$	yes

Table S7 | DenseNet-121, seed 0. Patient-clustered bootstrap intervals against the shape-matched control, computed as in Table S6. Six of ten regions remain significant after correction. The lung fields and mediastinum are negative; the spine, heart and diaphragmatic surface positive. Neither clavicle nor either scapula shows a statistically significant difference from its control.

Region	Δ	Uncorrected 95% CI	Bonferroni 99.5% CI	Sig.
Left lung field	-0.1343	$[-0.1526, -0.1158]$	$[-0.1608, -0.1078]$	yes
Mediastinum	-0.0342	$[-0.0449, -0.0235]$	$[-0.0496, -0.0193]$	yes
Right lung field	-0.0247	$[-0.0415, -0.0080]$	$[-0.0484, -0.0004]$	yes
Left clavicle	$+0.0031$	$[-0.0013,\ 0.0074]$	$[-0.0034,\ 0.0093]$	no
Right clavicle	$+0.0051$	$[\ 0.0000,\ 0.0105]$	$[-0.0025,\ 0.0131]$	no
Left scapula	$+0.0093$	$[-0.0011,\ 0.0198]$	$[-0.0059,\ 0.0244]$	no
Right scapula	$+0.0104$	$[-0.0007,\ 0.0212]$	$[-0.0056,\ 0.0257]$	no
Spine	$+0.0218$	$[\ 0.0106,\ 0.0330]$	$[\ 0.0061,\ 0.0381]$	yes
Heart	$+0.0556$	$[\ 0.0459,\ 0.0653]$	$[\ 0.0417,\ 0.0698]$	yes
Diaphragmatic surface	$+0.2449$	$[\ 0.2300,\ 0.2601]$	$[\ 0.2237,\ 0.2664]$	yes

Seed-level tests. One-sample t-tests of Δ against zero across three seeds carry two degrees of freedom and are severely underpowered; with the effect sizes observed under the shape-matched control they would be uninformative, and we do not report them. All inference rests on the patient-clustered bootstrap intervals above. Seed-level means and standard deviations appear in Table 5 of the main text.

S6 Comparison of ArcFace and Triplet Loss

In order to determine whether the regional ablation effects are dependent on the learning goal, we compared ArcFace with batch-hard triplet loss. Intact precision@1 performance was greater for ArcFace (0.9458) compared to triplet loss (0.8949). The regional ablation Δ's have high correlation ($r = 0.939$), including identical signs for seven out of ten regions (Table S8). However, the spine, right clavicle, and left scapula had differing signs. This table includes only ArcFace seed 0 from Table S6 and one triplet-loss trial instead of averaging across the three-seed ArcFace means shown in Table 5 of the main text.

Table S8 | Regional ablation effects under ArcFace and batch-hard triplet loss. The table shows Δ values at complete masking using ArcFace seed 0 and one triplet-loss trial. The results correlate highly ($r = 0.939$), with identical signs in seven out of ten regions.

Region	ArcFace Δ	Triplet Δ	Sign agrees
Right lung field	-0.0914	-0.1563	yes
Left lung field	-0.0789	-0.1269	yes
Mediastinum	-0.0453	-0.0051	yes
Spine	-0.0171	$+0.0902$	no
Right clavicle	-0.0013	$+0.0249$	no
Left clavicle	$+0.0153$	$+0.0298$	yes
Right scapula	$+0.0184$	$+0.0342$	yes
Left scapula	$+0.0204$	-0.0042	no
Heart	$+0.0520$	$+0.1278$	yes
Diaphragmatic surface	$+0.2836$	$+0.3472$	yes

S7 Compression check

To assess whether JPEG compression contributes to identification, we re-derived a subset of test images from the source PNGs without lossy compression and repeated cross-view retrieval against the same gallery. Precision@1 was 1.000 for both input formats across the 35 queries tested. This provides no evidence that compression artefacts contribute to identification, though the subsample is small.

S8 Spatial map of masking effects

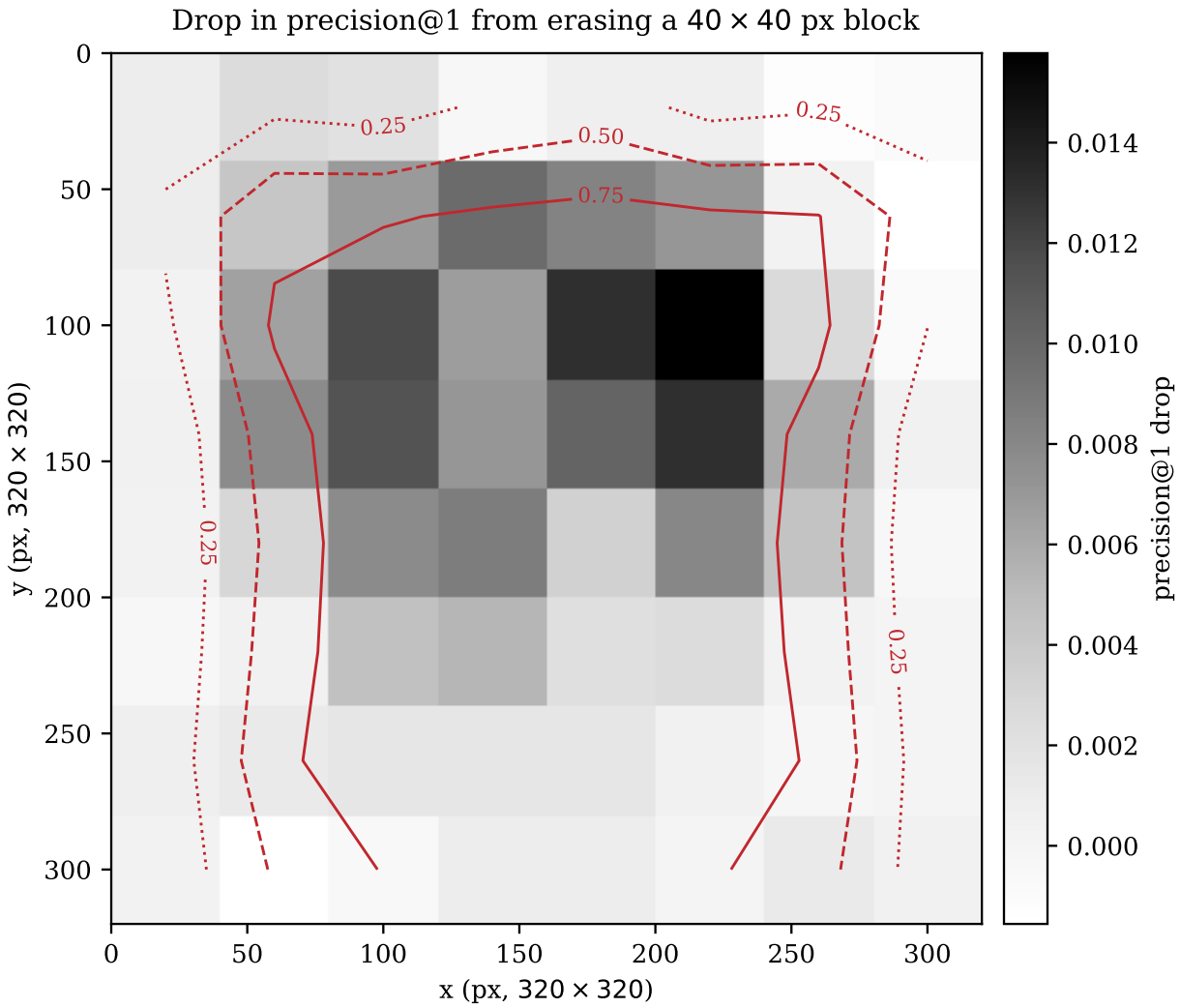


Figure S2 | Spatial sensitivity to masking (seed 0). Darker cells indicate larger drops in precision@1 after masking a 40×40 pixel block. The axes show mask position, while red contours indicate anatomical coverage. Retrieval loss generally increases with anatomical coverage, but varies substantially even at similar coverage levels ($R^2 = 0.38$). The largest drops occur in the upper-central thorax, suggesting that both anatomical coverage and mask location matter.

S9 Pairwise regional interactions

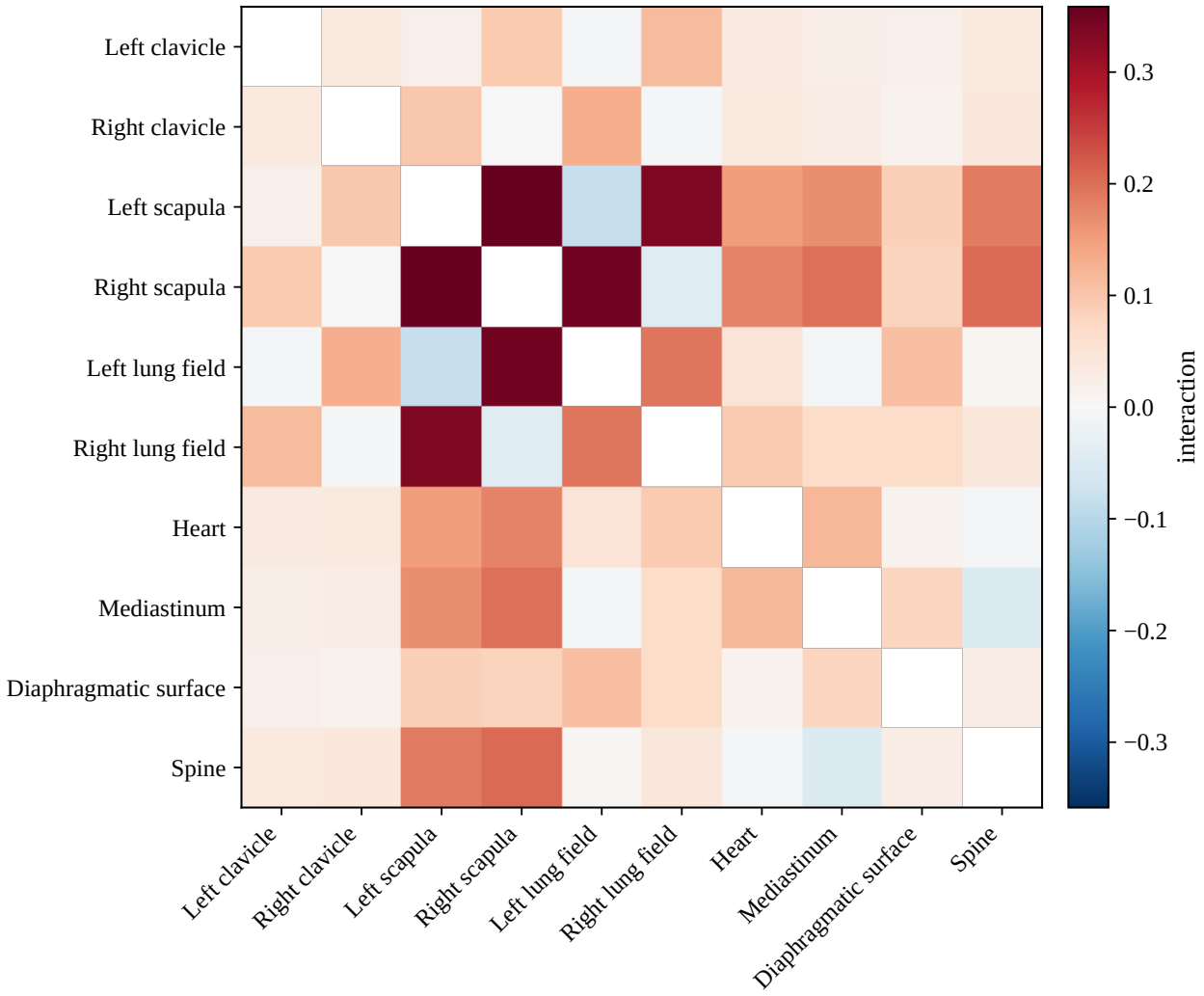


Figure S3 | Pairwise interactions in precision@1 loss. Each cell shows the combined loss from masking two regions minus the sum of their individual losses. Negative values indicate sub-additivity. All four same-side clavicle/scapula–lung pairs are negative, while all four opposite-side pairs are positive (seed 0).

S10 Sensitivity to the per-patient image cap

To examine sensitivity to the per-patient image cap, the ResNet-50 experiment was repeated with image caps of 10 and 5 images per patient, following the same training procedure (seed 0). The reference model for the cap-10 condition was re-trained for comparison and had intact precision@1 of 0.921 instead of the original model's value of 0.9458 seen in Table 2 of the main text. For all eight determinable regions, the regional Δ values showed signs consistent with those of the original model (Table S9).

Lowering the cap to 5 reduced intact precision@1 from 0.921 to 0.782 in the original test set,

and 0.759 in a separately re-capped test set. The regional Δ values were still strongly correlated across caps (Pearson $r = 0.98$; Spearman $\rho = 0.92$), and showed signs matching the original in eight out of nine determinable regions. The only region showing difference in Δ was the left lung field which went from $\Delta = -0.038$ to $\Delta = +0.016$. Re-capping of the test set preserved all effect directions with differences in Δ being less than 0.012.

Table S9 | Sensitivity of regional ablation effects to the per-patient image cap (ResNet-50, seed 0). Numbers represent Δ scores under complete regional masking for cap-10 and cap-5 models using the identical training process. Note that the cap-10 reference model is separately trained and different from the original model in Table 2 of the main text. Both original and independently capped test sets are used for cap-5 experiments.

Region	Cap 10 original test set	Cap 5 original test set	Cap 5 re-capped test set
Intact precision@1	0.9211	0.7820	0.7586
Left clavicle	+0.0142	+0.0438	+0.0369
Right clavicle	+0.0078	+0.0433	+0.0335
Left scapula	+0.0562	+0.0889	+0.0778
Right scapula	+0.0318	+0.0249	+0.0278
Left lung field	−0.0380	+0.0156	+0.0113
Right lung field	−0.0691	−0.0831	−0.0899
Heart	+0.0651	+0.0669	+0.0568
Mediastinum	−0.0280	−0.0147	−0.0188
Diaphragmatic surface	+0.3023	+0.2894	+0.2866
Spine	+0.0011	+0.0129	+0.0132