iApp Technology · Technical Whitepaper WP-2026-26

Image Background Remover: Cut-out Accuracy on Single-Subject Photographs

Production service as of 28 October 2026 · Measured 28 October 2026 · iApp Technology Co., Ltd., Bangkok, Thailand

Executive summary. We measured the iApp Image Background Remover on 802 photographs from the validation split of the public COCO image benchmark, each showing one clear subject whose outline was drawn by human annotators. On the 186 photographs of people, the service's cut-out overlapped the person with a mean intersection over union (IoU) of 0.790 (95% CI [0.760, 0.818]; median 0.857); 20 of the 186 were failed cut-outs (IoU below 0.5). Animals were cut out best (0.825), vehicles and other objects less well (0.722 and 0.651). People who fill the frame are harder than people seen at mid distance: 0.743 against 0.822. The public API returned the same cut-out as the production service for all 150 of 150 sampled images, with a median end-to-end time of 0.33 s.

1 Introduction

The iApp Image Background Remover takes a photograph and returns it as a PNG image whose transparency (alpha) channel removes the background, for ID photographs, profile pictures and product images. This paper asks how closely the cut-out follows the subject, on photographs the service has not seen, and where it fails. It reports every group we measured, including those where the service does poorly, and states the limits of the evidence in Section 7.

2 Test Data

The photographs come from the 2017 validation split of the COCO benchmark [1], whose object outlines were drawn by human annotators and are published under CC BY 4.0; each photograph keeps its own Flickr licence and is not redistributed here. A background remover is asked for one subject, so we kept the photographs that have exactly one clear subject:

- no crowd region is annotated in the photograph;
- the largest annotated object covers at least 10% of the frame;
- the second-largest object is at most a fifth of the subject's area, and all other objects together cover at most 3% of the frame outside the subject (objects worn or held inside the subject's outline are allowed);
- the subject is not a dining table or a bed, which are scene surfaces rather than subjects.

802 photographs qualify and all are used: 186 of people, 143 of animals, 218 of vehicles and 255 of other objects. They are consumer photographs, at most 640 pixels on the longer side (median 0.27 megapixels). The ground truth is the subject's annotated outline. The service's model was not trained on this collection.

3 Method

Each photograph was sent once, as the customer sends it, to the production service, which returned the cut-out as a PNG image. The alpha channel was scored against the subject's outline:

- **IoU**: the overlap between the kept area (alpha at least 50%) and the subject, divided by their union. 1 is a perfect cut-out.
- **Boundary F**: how many edge pixels of the cut-out lie within 0.8% of the image diagonal (about 7 pixels at 640 × 480) of the subject's outline, and the reverse, combined as an F-measure [2]. It measures the edge, where a cut-out is judged by eye.
- **MAE**: the mean absolute difference between the alpha channel (0 to 1) and the subject mask, over all pixels.
- **Failed cut-out**: IoU below 0.5. We also count near-empty results, where the service keeps less than 1% of the frame although the subject covers at least 10%.

Confidence intervals are seeded percentile bootstraps (2,000 resamples, seed 1337). 0 of 802 requests failed and 0 returned an image of a different size, so every photograph is scored.

4 Results

Table 1: Cut-out accuracy by subject. IoU with its 95% CI; "IoU ≥ 0.9" is the share of photographs cut out almost exactly; the last column counts failed cut-outs (IoU below 0.5).

Subject	Photos	Mean IoU [95% CI]	Median	IoU ≥ 0.9	Boundary F	MAE	Failed
People	186	**0.790** [0.760, 0.818]	0.857	33%	0.701	0.086	20
Animals	143	**0.825** [0.797, 0.851]	0.899	50%	0.754	0.064	10
Vehicles	218	**0.722** [0.680, 0.761]	0.859	39%	0.611	0.122	43
Other objects	255	**0.651** [0.610, 0.690]	0.782	35%	0.560	0.167	76
All images	802	**0.733** [0.714, 0.752]	0.856	39%	0.641	0.117	149

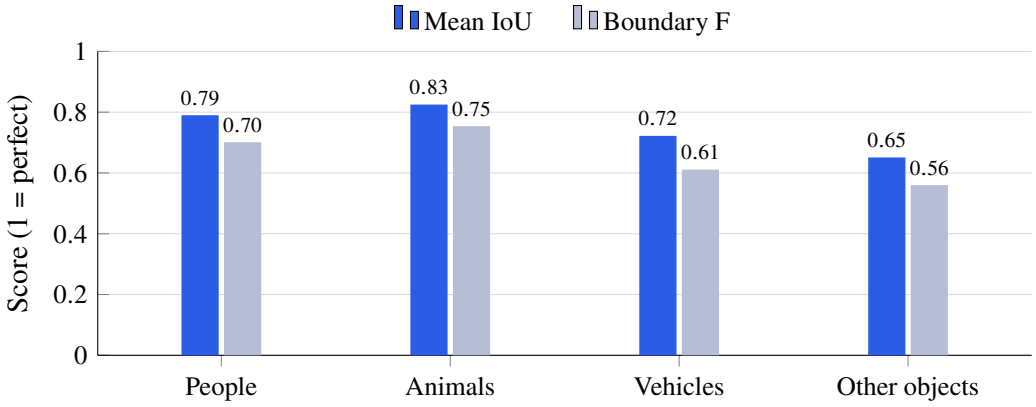


Figure 1: Mean IoU and boundary F by subject, 802 photographs.

Table 2: People, by the share of the frame the person fills.

Person fills	Photos	Mean IoU [95% CI]	Median	Boundary F	Failed
10–20%	83	0.815 [0.781, 0.844]	0.847	0.792	3
20–30%	34	0.822 [0.743, 0.877]	0.868	0.744	2
30% or more	69	0.743 [0.675, 0.802]	0.880	0.571	15

Table 3: Mean IoU for every animal, vehicle and object category with at least 10 photographs, weakest first (read down the left half, then the right).

Category	Photos	Mean IoU	Category	Photos	Mean IoU
bench	11	0.428	cat	23	0.789
oven	12	0.483	dog	18	0.800
refrigerator	16	0.498	pizza	16	0.802
vase	14	0.621	bird	11	0.808
boat	15	0.623	clock	12	0.815
teddy bear	13	0.645	giraffe	23	0.834
cake	10	0.651	zebra	16	0.835
train	83	0.677	elephant	10	0.842
airplane	19	0.683	bus	55	0.860
toilet	35	0.731	bear	22	0.867
truck	22	0.757	fire hydrant	13	0.877
motorcycle	12	0.772	stop sign	14	0.887
cow	10	0.775			

5 Discussion

Most cut-outs are good; the failures are total. The median IoU is well above the mean in every group (Table 1): most photographs are cut out closely, and a minority fail outright. Of the 149 failed cut-outs, 115 kept too little of the subject and 27 kept too much of the background; 15 returned an almost empty image.

People. The worst cut-outs of people, inspected one by one, follow three patterns. In dark or low-contrast scenes the service keeps almost nothing. When a person holds a striking object close to the camera, such as a cake or a telephone, the service cuts out the object and drops the person. Dark clothing against a dark background is lost while the face is kept. These cases are concentrated in close-ups, which is why people who fill at least 30% of the frame score lower (0.743, 15 of 69 failed) than people at mid distance (0.822; Table 2). The set holds few head-and-shoulders photographs against a plain background, the typical ID picture, so it does not measure that case on its own.

Objects. Animals are the strongest group (0.825). The weakest categories are large objects that belong to a room or a street scene, led by bench (0.428 over 11 photographs); compact, distinctive objects score highest, led by stop sign (0.887; Table 3). Among the 76 failed cut-outs of other objects, 52 kept too little of the object and 18 kept too much of the scene.

Edges. The boundary F of 0.641 over all photographs is lower than the IoU because the reference outlines are polygons drawn by annotators, not pixel-exact masks (Section 7); it is most useful for comparing groups.

6 Production Performance

A seeded sample of 150 photographs (seed 1337) was sent through the public API, `https://api.iapp.co.th/v3/store/smart-city/remove-background`, from Bangkok on 28 October 2026, two requests

at a time. All 150 succeeded, and all 150 returned an alpha channel bit-identical to the production service's: the scores above are the scores a customer receives.

Table 4: Time per photograph, seconds, measured at the client. Photographs are at most 640 pixels on the longer side.

Path	Photos	Median	90th pct.	95th pct.
Public API, end to end	150	0.33	0.41	0.43
Production service, inside iApp's network (3 at a time)	802	0.23	0.31	0.38

7 Limitations

- **Reference outlines are polygons.** They follow the subject within a few pixels but not hair, fur or fine edges. A pixel-accurate cut-out is therefore scored slightly below 1 on IoU and further below on boundary F.
- **One subject per photograph.** The selection favours clear single subjects. An object worn or held by a person is part of the person's outline only where the annotators drew it so; cutting out a second, unannotated object counts against the service.
- **Consumer photographs, not ID photographs.** The set has no studio portraits on plain backgrounds, the service's main use, and the photographs are at most 640 pixels on the longer side, smaller than a phone camera's.
- **One run.** The service is deterministic: the public API reproduced the production service's output exactly.

8 Reproducibility

The selection rule above is applied to the public COCO 2017 validation annotations by `get_data.py`; `bench.py` sends each photograph and stores the scores, and `gen_results.py` produces every figure in this paper. The per-photograph scores and the scripts are available from iApp Technology on request: sale@iapp.co.th. The public API calls in this paper used 151 IC.

References

[1] T.-Y. Lin, M. Maire, S. Belongie et al. Microsoft COCO: Common Objects in Context. *ECCV*, 2014. Annotations CC BY 4.0. https://cocodataset.org

[2] F. Perazzi, J. Pont-Tuset, B. McWilliams et al. A Benchmark Dataset and Evaluation Methodology for Video Object Segmentation. *CVPR*, 2016.

About iApp Technology. iApp Technology Co., Ltd. is a Thai artificial-intelligence company providing OCR, document understanding, speech, identity verification and large-language-model services for the Thai language. Web: iapp.co.th · E-mail: sale@iapp.co.th · Tel: +66 2 124 4041