



Passport MRZ OCR: A Self-Validating Accuracy Benchmark

Service version v2.0.3-20260825 · Measured 25 August 2026 · iApp Technology Co., Ltd., Bangkok, Thailand

Executive summary. A passport's machine-readable zone is self-validating: under ICAO 9303 every key field carries a check digit computed from its own contents, so the correctness of an OCR read can be measured arithmetically, with no reference transcription and no judgement calls. On a 266-passport benchmark spanning many issuing countries and capture conditions, the iApp Passport OCR service (v2.0.3-20260825) returns the document number, date of birth, and expiry date **all verified against their printed check digits on 95.5%** of passports, up from 58.6% in the previous release, with **100%** of MRZs parsed and a median response time of 0.2 s.

1 Introduction

The iApp Passport OCR API reads the machine-readable zone of any ICAO 9303 passport and returns the holder's details, the face photograph, and a set of verification flags. It is operated by iApp Technology Co., Ltd. as a public cloud API at api.iapp.co.th and as a fully self-contained on-premise deployment.

Most OCR benchmarks depend on a reference transcription, whose own quality bounds what the benchmark can claim. Passport MRZs permit something stronger: each of the document number, date of birth, expiry date, and personal number carries a check digit printed on the passport itself, and the line as a whole carries a composite check digit. A returned field that disagrees with its own check digit is wrong, arithmetically — a document number that fails its check digit cannot belong to any real passport.

2 Method

Every figure in this paper is the share of the 266 benchmark passports whose returned field satisfies the check digit printed on that passport.

- **Test set.** 266 passports spanning many issuing countries, image qualities, and capture angles, including deliberately rotated and augmented samples. The set contains real identity documents and is withheld under Thai PDPA obligations.
- **A conservative measure.** A check digit is a single decimal digit, so roughly one wrong read in ten satisfies it by coincidence and is scored as correct. True accuracy is therefore slightly below the reported figures — in both releases equally, so the comparison stands.
- **No self-grading.** The release adds a structural validation layer that only emits repaired fields whose check digit passes; measuring that layer by check digits would be circular. The layer is therefore evidenced separately (Section 4), and the engine comparison is run with it disabled.

3 Results

In practical terms: the previous release returned a document number that could not belong to any real passport on 30% of passports; v2.0.3 reduces that to under 4%, and flags the residual cases through the `valid_number` field so integrators can trigger a re-capture instead of storing a wrong number.

Measure	Previous release	v2.0.3-20260825
MRZ parsed	98.5%	100.0%
Document number verified	69.9%	96.2%
Date of birth verified	84.6%	99.2%
Expiry date verified	82.3%	98.1%
Personal number verified	73.3%	98.5%
Composite check digit verified	56.4%	95.5%
Number, birth date, and expiry all verified	58.6%	95.5%
Median response time	0.26 s	0.19 s

Table 1: Share of the 266 benchmark passports whose returned fields satisfy their printed check digits.

4 The Two Improvements, Evidenced Separately

The release combines an upgraded recognition engine with a structural validation layer. Each is measured on its own evidence.

4.1 Recognition engine (validation layer disabled)

Check digit	Previous engine	New engine alone
Document number	69.9%	95.1%
Date of birth	84.6%	98.5%
Expiry date	82.3%	97.4%
Personal number	73.3%	85.7%
All three primary fields	58.6%	94.4%

Table 2: Engine-only comparison on identical images, structural validation disabled.

4.2 Structural validation

An MRZ is a fixed 2×44-character grid in which every position has a known character class. The validation layer normalises the recognised lines to that grid, enforces the character class per position, and repairs a field that fails its check digit only when substituting a single character from a known visual-confusion pair produces exactly one reading that satisfies the digit.

It is deliberately conservative: it never rewrites a check digit to match a field, never changes more than one character, and declines to act when more than one candidate reading would pass. Under synthetic single-character corruption of known-good MRZs it recovers 100% of corrupted dates and 82% of fields overall, and — the property that matters most — of 774 fields that already satisfied their check digits before the layer ran, **it altered none**.

5 Production Performance

6 Deployment and Data Security

The public cloud API is GDPR- and PDPA-compliant: uploaded images are processed in memory and are not retained after the response is returned. The service runs entirely on-premise-capable infrastructure —

Metric	Value
Median end-to-end processing time	0.2 s per passport
Sustained throughput	2.6 passports/s ($\approx 9,500$ passports/hour)
Supported inputs	JPEG, PNG, WEBP, PDF (per-page results)
Coverage	Any ICAO 9303 machine-readable passport

Table 3: Production service performance, measured on the deployed v2.0.3-20260825 service, August 2026.

no image is sent to any external service — so a fully self-contained deployment is available in which no passport data leaves the customer’s infrastructure.

7 Reproducibility

The measurement requires no proprietary ground truth: any party can score this service, or any competitor, by submitting their own passports and checking the returned fields against the check digits printed on the documents — the method in Section 2 is complete. The benchmark set itself is withheld because it contains real identity documents. Enquiries may be addressed to sale@iapp.co.th.

About iApp Technology. iApp Technology Co., Ltd. is a Thai artificial-intelligence company providing OCR, document understanding, speech, and large-language-model services for the Thai language.

Web: iapp.co.th · E-mail: sale@iapp.co.th · Tel: +66 86 322 5858