Online Resource 1: OIDIQ: Open identity document image quality

Gregor Grote [1*], Juan E. Tapia [1] and Christian Rathgeb [1]

[1*] da/sec, Hochschule Darmstadt, Schöfferstraße 3, Darmstadt, 64295, Hessen, Germany.

*Corresponding author(s). E-mail(s): gregor.grote@h-da.de;
Contributing authors: juan.tapia-farias@h-da.de; christian.rathgeb@h-da.de;

Online Resource 1 of the article "OIDIQ: Open identity document image quality", International Journal on Document Analysis and Recognition (IJDAR).

This document gives the scalar conversion of each quality measure of the paper (Section 5.4), in the order of Section 5. Numbers with the prefix S refer to this document, all others to the paper. All scores are clipped to $[0, 100]$.

S1 Minimum corner confidence

The error stays small at high confidence and increases quickly below a certain confidence. We therefore fitted the maximum corner error e against the minimum corner confidence c with the curve

$$\hat{e}(c) = k\,(1 - c), \tag{S1}$$

by minimizing the squared difference of the logarithms of e and $\hat{e}$, with equal weights for the test sets (Figure S1). The fit on all test sets gives $k = 11.83$. This curve fits best of the simple curves we compared on left-out test sets. In an operational scenario this function might vary. The factor is similar across the test sets (k between 10.60 and 13.24), and the fit has an R^2 of 0.59 on the logarithmic scale.

The score is:

$$\begin{aligned} S_{\text{corner}} &= 100 \cdot \frac{e_{\text{low}} - \hat{e}(c)}{e_{\text{low}} - e_{\text{high}}} \\ &= 100 \cdot \frac{k(1 - c_{\text{low}}) - k(1 - c)}{k(1 - c_{\text{low}}) - k(1 - c_{\text{high}})} \\ &= 100 \cdot \frac{c - c_{\text{low}}}{c_{\text{high}} - c_{\text{low}}}, \end{aligned} \tag{S2}$$

where e_{high} is the error for a score of 100 and e_{low} the larger error for a score of 0. The confidences $c_{\text{high}} = 1 - e_{\text{high}}/k$ and $c_{\text{low}} = 1 - e_{\text{low}}/k$ are the points where the fit reaches these errors. As $\hat{e}$ is linear in c, the score is also linear in c, and k only sets where the two points lie. We chose $e_{\text{high}} = 1\,\%$ and $e_{\text{low}} = 5\,\%$ of the ID diagonal as an example. These errors are about 12 and 59 pixels on the normalized image. They correspond to confidences in the fit of $c_{\text{high}} = 0.92$ and $c_{\text{low}} = 0.58$. With equal weights for the test sets in the fit, 73.1 % of the IDs reach a score of 100, with a median maximum corner error of 0.63 % and a 95th percentile of 1.42 %. A share of 2.0 % of the IDs gets a score of 0, with a median maximum corner error of 7.43 %.

S2 ID confidence

We based the score on the share p of wrong detections (Section 5.1), which we fitted against the ID

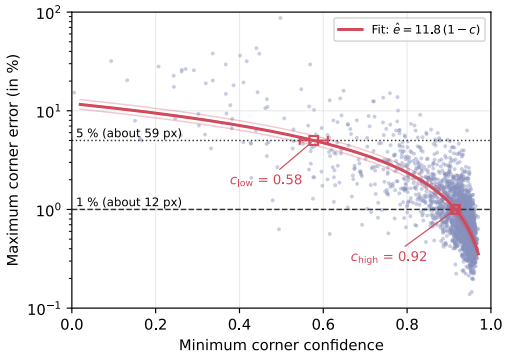


Fig. S1 Maximum corner error of an ID in % of the ID diagonal against its minimum corner confidence on the five test sets with IDs, with the fit per test set (thin) and on all IDs with equal weights for the test sets (bold), and the two anchor points of the score with their standard deviation over the fits. Of the 19,936 IDs, 600 per test set are shown

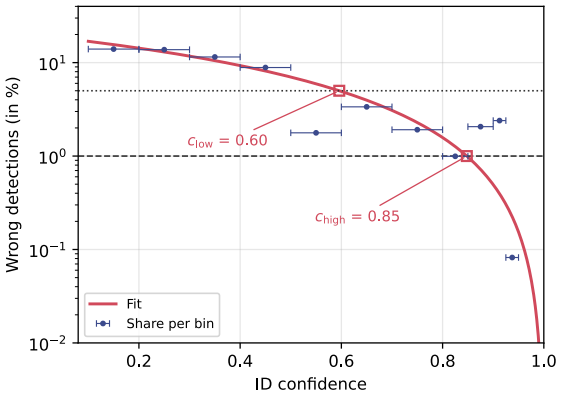


Fig. S2 Share of wrong detections against the ID confidence, with the fit and the two anchor points of the score. Dots: share per confidence bin, horizontal lines: width of the bin. Bins without wrong detections lie below the axis

confidence c (Figure S2):

$$\ln \frac{p}{1-p} = a + b \ln(1-c). \tag{S3}$$

We evaluated several simple curves with the leave-one-out approach on the test sets, and selected the one that performed best overall. The fit on all test sets, with equal weights for the test sets, gives $a = -1.41$ and $b = 1.69$, with bootstrap 95 % confidence intervals of -1.74 to -1.09 and 1.50 to 1.96.

The score is:

$$\begin{aligned} S_{\text{ID}} &= 100 \cdot \frac{p_{\text{low}} - \hat{p}(c)}{p_{\text{low}} - p_{\text{high}}}, \\ \hat{p}(c) &= \frac{1}{1 + e^{-a}(1-c)^{-b}}, \end{aligned} \tag{S4}$$

where p_{high} is the share for a score of 100 and p_{low} the larger share for a score of 0. Solving the fit for c gives the confidence at which it reaches a share p:

$$c(p) = 1 - \exp\left(\frac{1}{b}\left(\ln \frac{p}{1-p} - a\right)\right), \tag{S5}$$

so that the score is 100 from $c_{\text{high}} = c(p_{\text{high}})$ and 0 up to $c_{\text{low}} = c(p_{\text{low}})$. We chose $p_{\text{high}} = 1\,\%$ and $p_{\text{low}} = 5\,\%$. They correspond to $c_{\text{high}} = 0.85$ (bootstrap 95 % confidence interval 0.81 to 0.87) and $c_{\text{low}} = 0.60$ (0.52 to 0.65). With the curve fitted to all test data, 92.7 % of the detections reach a score of 100, of which 0.17 % are wrong, and 3.2 % get a score of 0, of which 10.28 % are wrong.

S3 Occlusion mask confidence

We fitted the error $e = 1 - \text{IoU}$ against the confidence with the curve

$$\hat{e}(c) = k\,(1-c), \tag{S6}$$

by minimizing the squared difference of the logarithms of e and $\hat{e}$, with equal weights for the test sets (Figure S3). The fit on both test sets gives $k = 2.94$ (bootstrap 95 % confidence interval 2.81 to 3.08). The factor is similar across the test sets (2.84 to 3.04), and the fit has an R^2 of 0.59 on the logarithmic scale.

The score is:

$$\begin{aligned} S_{\text{mask}} &= 100 \cdot \frac{e_{\text{low}} - \hat{e}(c)}{e_{\text{low}} - e_{\text{high}}} \\ &= 100 \cdot \frac{k(1-c_{\text{low}}) - k(1-c)}{k(1-c_{\text{low}}) - k(1-c_{\text{high}})} \\ &= 100 \cdot \frac{c - c_{\text{low}}}{c_{\text{high}} - c_{\text{low}}}, \end{aligned} \tag{S7}$$

where e_{high} is the error for a score of 100 and e_{low} the larger error for a score of 0. The confidences $c_{\text{high}} = 1 - e_{\text{high}}/k$ and $c_{\text{low}} = 1 - e_{\text{low}}/k$ are the points where the fit reaches these errors. As $\hat{e}$ is linear in c, the score is also linear in c, and k only sets where the two points lie. We chose

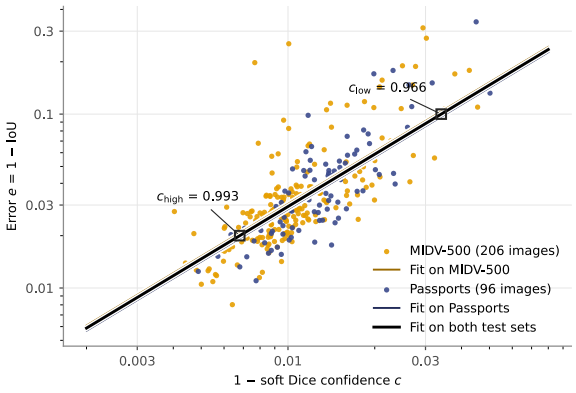


Fig. S3 Error of the occlusion mask against the soft Dice confidence on both test sets, with the fit per test set (thin) and on all images with equal weights for the test sets (bold), and the two anchor points of the score. Both axes are logarithmic

$e_{\text{high}} = 0.02$ and $e_{\text{low}} = 0.10$, that is, an IoU of 0.98 and 0.90.They correspond to confidences of $c_{\text{high}} = 0.993$ and $c_{\text{low}} = 0.966$.

S4 Illumination artifacts

We set the score of the maximum x of the predicted map (Section 5.2.1) by the false positive rate (FPR), the share of test images without the artifact whose maximum reaches a threshold (Figure S4). A score of 100 marks the threshold x_{20} at which 20 % of these images are flagged, and a score of 0 the threshold x_0, the largest maximum of these images, at which none of them are flagged. The tolerable FPR may differ from operation to operation, and the thresholds may vary with the dataset and the capture conditions. The score is:

$$S_{\text{illumination}} = 100 \cdot \frac{x_0 - x}{x_0 - x_{20}}, \tag{S8}$$

where $S_{\text{illumination}}$ is rounded to the nearest integer and limited to the range from 0 to 100. For glares, this gives $x_{20} = 0.12$ and $x_0 = 0.88$ (bootstrap 95 % confidence intervals 0.08 to 0.17 and 0.72 to 0.88), and 91 % of the images with glares score below 100 and 36 % score 0. For shadows, this gives $x_{20} = 0.09$ and $x_0 = 0.25$ (0.06 to 0.13 and 0.16 to 0.25), and all images with shadows score 0.

S5 Compression artifacts

The score of the estimated quality $\hat{q}$ (Section 5.2.2) is:

$$S_{\text{compression}} = 100 \cdot \frac{\hat{q}}{q_t}. \tag{S9}$$

As the original images may already be compressed, we do not penalize slight compression and use $q_t = 80$, so that an image with $\hat{q} \geq 80$ scores 100. As $\hat{q}$ also reflects the unknown compression of the original, we tested the score by its change from the original to the compressed image, with the curve fitted on the other dataset (Figure S5). For JPEG and JPEG 2000, the change follows q with a Spearman correlation between 0.90 and 0.93, and for WebP with 0.70 on FantasyID and 0.68 on MIDV-500.

Figure S6 shows how these correlations depend on q_t. For JPEG, the correlation peaks at about $q_t = 90$ on FantasyID and at about $q_t = 70$ on MIDV-500, and $q_t = 80$ lies between them.

S6 Sharpness

To convert the output x of our sharpness model to a scalar, we fitted multiple simple curves to the CER of the OCR example (Section 5.2.3) against x on all images (Figure S7). Compared on left-out document types, the logistic curve fits best:

$$\widehat{\text{CER}}(x) = \frac{A}{1 + e^{(x - x_0)/s}}. \tag{S10}$$

The fit on all images gives $A = 1.052$, $x_0 = 0.522$ and $s = 0.098$. The parameters are similar across the folds (A between 1.039 and 1.077, x_0 between 0.517 and 0.525, s between 0.095 and 0.103). The score is:

$$S_{\text{sharpness}} = 100 \cdot \frac{\text{CER}_{\text{low}} - \widehat{\text{CER}}(x)}{\text{CER}_{\text{low}} - \text{CER}_{\text{high}}}, \tag{S11}$$

where CER_{high} is the fitted CER that gives the highest quality score, $S_{\text{sharpness}} = 100$, and CER_{low} the fitted CER that gives the lowest, $S_{\text{sharpness}} = 0$. We chose $\text{CER}_{\text{high}} = 5\,\%$ and $\text{CER}_{\text{low}} = 75\,\%$ as an example. They correspond to $x = 0.815$ and $x = 0.433$ (bootstrap 95 % confidence intervals 0.803 to 0.829 and 0.428 to 0.438).

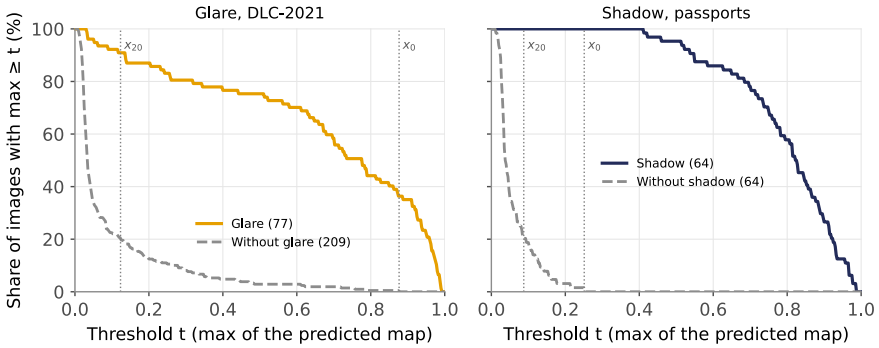
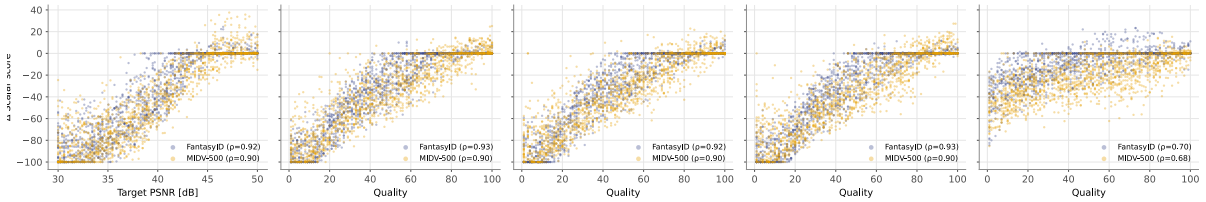


Fig. S4 Share of the test images with and without the artifact whose maximum of the predicted map reaches the threshold t, for glares on DLC-2021 (left) and shadows on the passports (right). The dotted lines mark the thresholds x_{20} (score 100) and x_0 (score 0) of the scalar conversion



(a) JPEG 2000 (b) JPEG 4:2:0 (c) JPEG 4:2:2 (d) JPEG 4:4:4 (e) WebP

Fig. S5 Change of the compression score $S_{\text{compression}}$ ($q_t = 80$, curve fitted on the other dataset) between the original and the compressed image against the quality parameter q (for JPEG 2000: the target PSNR in dB) for each compression setting, with the Spearman correlation ρ per dataset. For readability, each plot shows a random subset of 1,500 images per dataset, while ρ is computed on all images

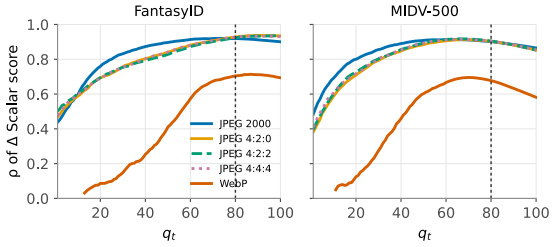


Fig. S6 Spearman correlation ρ between the change of the compression score and q against q_t for each compression setting, with the curve fitted on the other dataset. The dotted line marks $q_t = 80$. The WebP curves start later, as all their scores are 100 for very low q_t

S7 Over- and underexposure

We set the scores of the mean gray value m (Section 5.2.4) by the false positive rate (FPR) on our dataset. A score of 0 marks the threshold m_1 at which 1 % of these frames are flagged and a score of 100 the threshold m_{20} at 20 %, so that 80 % of the frames without the label score 100. The tolerable FPR may differ from operation to operation. For underexposure, the score is:

$$S_{\text{underexposure}} = 100 \cdot \frac{m - m_1}{m_{20} - m_1}. \qquad \text{(S12)}$$

The score $S_{\text{overexposure}}$ is the mirror image, with the thresholds above the gray values of the clean frames. This gives $m_1 = 121.9$ and $m_{20} = 157.0$ for underexposure and $m_1 = 227.5$ and $m_{20} = 190.4$ for overexposure. The bootstrap 95 % confidence intervals of the thresholds are at most 10.5 gray values wide.

A score below 100 flags 93 % of the underexposed and 85 % of the overexposed frames, and a score of 0 flags 7 % and 46 %. To test whether the FPR holds on unseen documents, we took the thresholds from four folds of the document types and scored the fifth. A score below 100 then flags 20.6 % of the frames without the label for underexposure and 20.7 % for overexposure, and a score of 0 flags 1.7 % and 1.2 %, close to the rates we set.

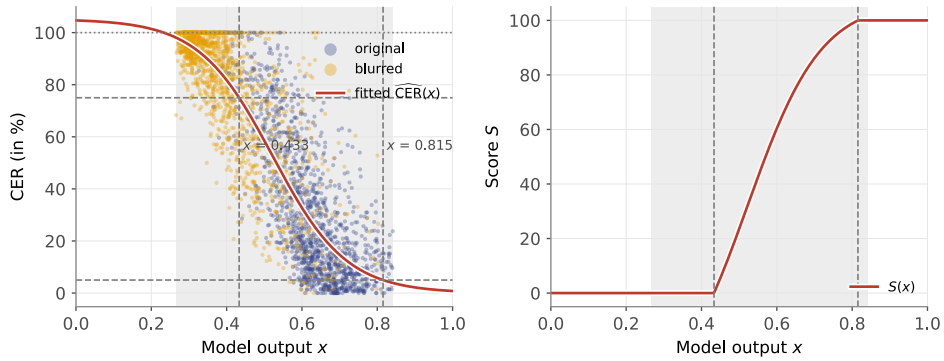


Fig. S7 Scalar conversion of the sharpness score over all model outputs x: (a) fitted CER $\widehat{\mathrm{CER}}(x)$ (red) with the OCR example images (1,500 of the 2,000 original images in blue and 1,500 of the 2,000 blurred images in yellow), and (b) the resulting score $S(x)$. The gray band marks the range of x on these images, and the dashed lines mark $\mathrm{CER}_{\mathrm{high}} = 5\,\%$ at $x = 0.815$ and $\mathrm{CER}_{\mathrm{low}} = 75\,\%$ at $x = 0.433$

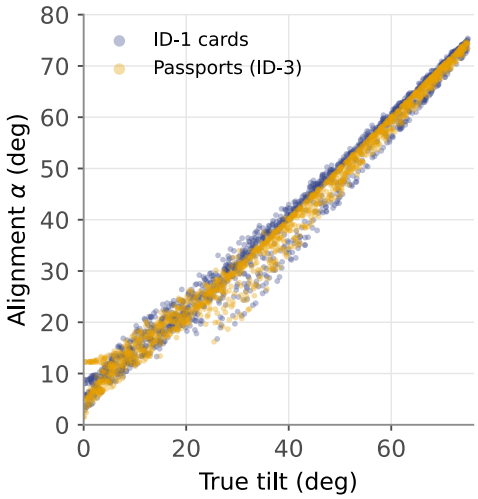


Fig. S8 Mean tilt α against the true tilt for 1,500 simulated ID-1 cards (blue) and 1,500 simulated passports (yellow), without corner noise

S8 Alignment

The score of the mean tilt α (Section 5.3.1) is:

$$S_{\mathrm{alignment}} = 100 \cdot \frac{\alpha_{\max} - \alpha}{\alpha_{\max} - \alpha_{\min}}. \tag{S13}$$

We chose $\alpha_{\min} = 0°$ and $\alpha_{\max} = 30°$. Figure S8 shows α against the true tilt of the simulated quadrilaterals (Section 5.3.1).

S9 ID size

The score of the relative size r (Section 5.3.2) has two pieces, on both sides of the ideal size:

$$S_{\mathrm{size}} = \begin{cases} 100 \cdot \dfrac{r - r_{\min}}{r_{\mathrm{ideal}} - r_{\min}} & \text{if } r \le r_{\mathrm{ideal}}, \\ 100 \cdot \dfrac{r_{\max} - r}{r_{\max} - r_{\mathrm{ideal}}} & \text{if } r > r_{\mathrm{ideal}}. \end{cases} \tag{S14}$$

We chose $r_{\mathrm{ideal}} = 85\,\%$, $r_{\min} = 50\,\%$ and $r_{\max} = 100\,\%$.Above $r_{\max}$, the score is 0.

S10 ID in frame

For the share x of the ID outside the image (Section 5.3.3), we tolerate 0.1 %, as a small corner error can push the quadrilateral of an ID close to the border over it. The score is:

$$S_{\mathrm{frame}} = 100 \cdot \frac{x_{\max} - x}{x_{\max} - x_{\min}}, \tag{S15}$$

with $x_{\min} = 0.1\,\%$ and $x_{\max} = 3\,\%$.

S11 Occlusion

We use the conversion of the face occlusion measure of OFIQ [1] for the occluded share o (Section 5.3.4):

$$S_{\mathrm{occlusion}} = 100\,(1 - w\,o)\,, \tag{S16}$$

where w is a factor that sets how strongly occlusions are penalized. OFIQ uses $w = 1$, meaning

that if 20% of the face is occluded, the scalar score would still be 80. We propose $w = 10$, so that if 5% of the ID is occluded, the scalar score would be 50.

S12 Only one ID present

We use the conversion of the single face present measure of OFIQ [1]:

$$S_{\text{single}} = 100 \cdot \left(1 - \frac{a_2}{a_1}\right), \tag{S17}$$

with the areas a_1 and a_2 of the two largest detected IDs (Section 5.3.5). The score is thus 100 for a single ID and 0 for two IDs of the same size. If no ID is detected, the score is defined to be 0.

References

[1] Open source face image quality (ofiq) – projektabschlussbericht. Project final report, Bundesamt für Sicherheit in der Informationstechnik (BSI) (july 2024). Reference implementation for ISO/IEC 29794-5. https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/OFIQ/Projektabschlussbericht_OFIQ_1_0.pdf