

000
001
002
003
004
005
006
007
008
009
010
011
012
013
014
015
016
017
018
019
020
021
022
023
024
025
026
027
028
029
030
031
032
033
034
035
036
037
038
039
040
041
042
043
044
045
046
047
048
049
050
051
052
053

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

A Deep Adversarial Framework for Visually Explainable Periocular Recognition

Anonymous CVPRW 2021 submission

Paper ID ****

Abstract

The ability to portray the reasoning behind a decision has been at the core of major research efforts. It serves not only to increase trust amongst the stakeholders of the automated agent, but also to potentially improve the entire system as a whole. In this work, we present our efforts towards a visually explainable periocular recognition framework, with a simple, yet effective solution that automatically provides a visual representation of the features in each region that sustained an impostor pairwise comparison. Based in our quantitative and qualitative experiments, the results validate the proposed goals and reiterate the notion that explainability should be strongly considered when designing ML algorithms.

1. Introduction

This work describes an integrated framework for periocular biometric recognition which - singularly - also provides the visual explanation that sustained each *impostor* pairwise comparison decision. In this context, we consider not only the recognition accuracy, but also the ability of the framework to portray the reasoning that supports every non-match decision. The latter is becoming an integral part of Machine Learning systems, given how ubiquitous and dependable they have become in recent years [2]. Thus, we diverge from the black-box paradigm and embrace a *visually explainable nature*, as illustrated in Fig. 1.

Typically, a recognition task involves a set of unique and non-transferable features that, when given to a system designed to do so, can unmistakably identify a subject. Biometrics, as they are designated in the field, serve such purposes, as long as they are universal, distinguishable, resilient to changes and easy to collect [1]. Upon proving compliant with the aforementioned requirements, biometrics can be separated into two dominant categories:

⁰The code is publicly available at <http://github.com/anonymized>

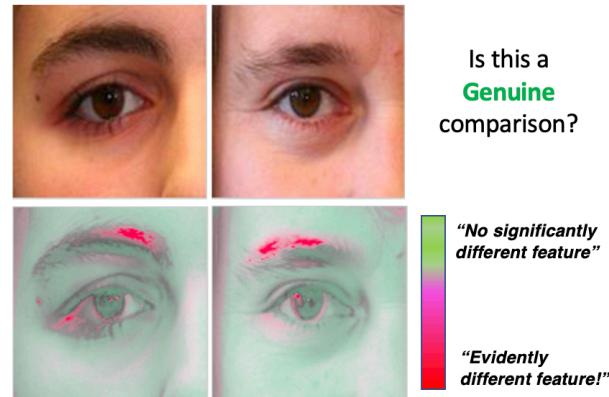


Figure 1. Key insight the proposed visual explanation framework: given a pair of samples, the system **not only reports a binary decision** (*genuine/impostor* classes), but also **highlights the regions in each sample that contributed the most** in case of a **non-match decision**. In this example, yet the iris and skin colour are similar between samples, the eyebrows and eyelashes shapes are evidently different, along with a skin spot in the left side sample.

1) *physiological* features, as the irises, fingerprints and retinas are naturally possessed by a given subject; and 2) *behavioural* biometrics, manifested when a subject interacts with the surrounding environment (e.g., the gait and handwritten signature traits) [18].

As a topic of growing interest in the biometrics domain, periocular recognition uses the rich area comprising and surrounding the eye, in which the iris, sclera, eyebrow, eyelid and skin stand out. We also acknowledge this set of facial components, considering, where applicable, their colour and/or shape.

Regarding explainability and its application to recognition problems, it is important to keep in mind that Deep Learning solutions rely on model complexity and abstraction prowess to become truly accurate. Although seemingly innocuous, there could be seriously negative outcomes if opaque algorithms gamble on the clearance of unauthorised people into sensible areas. It becomes clear, then, that including explainable components into AI

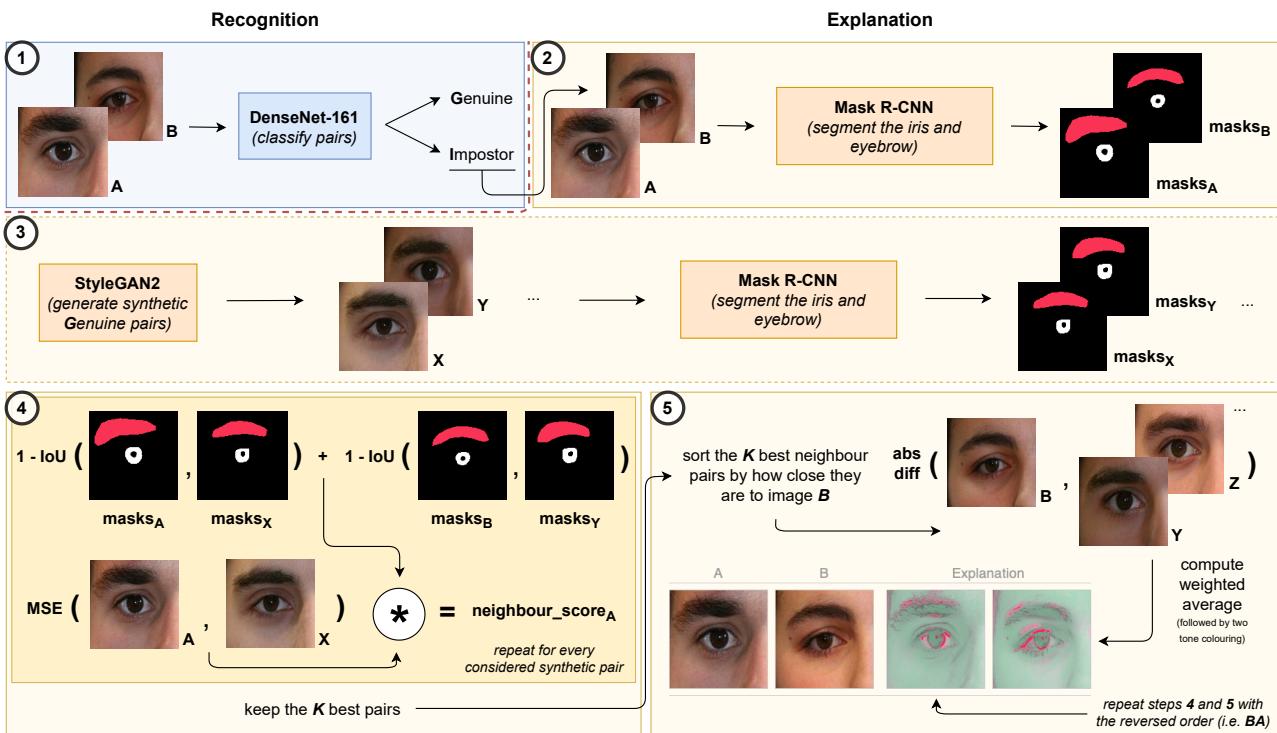


Figure 2. Diagram of the main pipeline. Step one (i.e. recognition) encompasses a CNN that distinguishes between *genuine* and *impostor* pairs. Then, if deemed *impostor*, steps two to five (i.e. explanation) try to find *genuine* synthetic pairs that closely resemble the test pair. By doing so, and despite looking similar, the test pair will probably contain certain internal differences (i.e. between images *A* and *B*) that the synthetic ones do not, thus providing an interpretable explanation.

systems is imperative. More recently, the politicians have addressed this urgency in formal terms. Recently, the EU, through the GDPR[5], introduced the notion of "*right to an explanation*". The definition and scope are still subject to debate [20], but these are definite strides towards an objective regulation regarding the explainable depth of autonomous systems.

According to the above paragraphs, this paper describes a framework that receives a pair of images (either *genuine* or *impostor*) and produces a two-fold output: a binary match/non-match decision and a *visual explanation* that highlights the features/regions of the input data that sustained the most a particular decision. This is seen as the main contribution of our work, in the sense that - to the best of our knowledge - it is the first that creates an accurate and explainable representation of the reasons behind each decision of the recognition system. Other important insights include the use of the powerful generative capabilities of Generative Adversarial Networks (GANs), to create samples that wouldn't otherwise exist in the training data, thus augmenting the variety and flexibility of the learning set (this process can also be seen as a form of data augmentation).

Fig. 2 provides a cohesive overview of the framework that performs the periocular recognition task and its subsequent explanation: 1) a CNN (of a well known architecture) is initially trained to optimally produce a match/non-match decision. If the pair is deemed to be an *impostor* comparison, a search process finds the most similar looking *genuine* pairs. The key here is that, even if the test pair has significant disparities between images *A* and *B* that led to an *impostor* decision, the closest synthetic pairs most likely do not (as they were all drawn from the *genuine* distribution). Then, by obtaining the pixel-wise weighted differences between the test pair and its *K* closest neighbours, the data disparities become evident. This assumption is justified by the CNN's non-match decision, which can likely be attributed to differences in iris colour, skin texture and color, or eyebrows and eyelids shapes/distributions.

The remainder of this paper is organised as follows: Section 2 summarises the most relevant research in the fields of periocular recognition and Machine Learning Explainability; Section 3 describes our method; Section 4 analyses the main results and Section 5 concludes this paper, providing some final remarks.

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215

| | | |
|-----|--|-----|
| 216 | 2. Related Work | 270 |
| 217 | | 271 |
| 218 | 2.1. Periocular Recognition | 272 |
| 219 | | 273 |
| 220 | The seminal breakthroughs in periocular recognition | 274 |
| 221 | tasks can be traced to a set of methods termed <i>feature de-</i> | 275 |
| 222 | <i>scriptors</i> . Methods such as HoG, LBP and SIFT were | 276 |
| 223 | able to produce simplified representations by relying on | 277 |
| 224 | edges, textures and keypoints, respectively. In [22], the | 278 |
| 225 | results from each feature descriptor are fused to provide | 279 |
| 226 | a more comprehensive description of an image’s content. | 280 |
| 227 | This work served as basis for consequent fusion based | 281 |
| 228 | approaches, as in [6]. Extending towards the field of super- | 282 |
| 229 | vised learning, in [11] a Restricted Boltzmann Machine is | 283 |
| 230 | used to learn a probabilistic distribution over the input data, | 284 |
| 231 | further discriminated with metric learning and SVMs. | 285 |
| 232 | With the effective application of Deep Learning solu- | 286 |
| 233 | tions, researchers turned to popular architectures, like Con- | 287 |
| 234 | volutional neural Networks (CNNs), in the pursuit of ever | 288 |
| 235 | increasing recognition accuracy. Accordingly, in [23] the | 289 |
| 236 | main concept involves the use of multiple CNNs that are | 290 |
| 237 | trained to become specialised in classifying certain seman- | 291 |
| 238 | tic information (e.g. gender, age, and more). Then, a score | 292 |
| 239 | fusion process creates a unified architecture. In [17], the | 293 |
| 240 | authors enforce a CNN to ignore the ocular region (due to | 294 |
| 241 | this region’s tendency to capture light reflections and other | 295 |
| 242 | performance degradation factors) and rely more on the sur- | 296 |
| 243 | rounding area (eyebrow, eyelid and skin). Going against | 297 |
| 244 | the idea of only relying on either the ocular or periocu- | 298 |
| 245 | lar regions, in [19] the iris and periocular biometrics are | 299 |
| 246 | separately explored for classification purposes, with the | 300 |
| 247 | resulting scores being fused to reach a final decision. More | 301 |
| 248 | recently, in [4] the authors bridge the gap between biomet- | 302 |
| 249 | ric recognition (in their case, facial) and interpretability, by | 303 |
| 250 | learning feature specific filters that activate in a range of | 304 |
| 251 | preferred spatial locations, and, in [9], an integrated solu- | 305 |
| 252 | tion is proposed by leveraging part discovery as a form of | 306 |
| 253 | attention. However, a fusion of periocular recognition and | 307 |
| 254 | interpretability remains largely unexplored, thus motivating | 308 |
| 255 | the development of novel solutions. | 309 |
| 256 | | 310 |
| 257 | | 311 |
| 258 | 2.2. Machine Learning Explainability | 312 |
| 259 | | 313 |
| 260 | In the literature, explainable techniques are commonly | 314 |
| 261 | discriminated in terms of depth, scope and model applica- | 315 |
| 262 | bility [12], [15]. Depth is related to the length to which we | 316 |
| 263 | explain a given model, i.e. whether the technique limits the | 317 |
| 264 | model’s complexity to make it more transparent (<i>intrinsic</i>) | 318 |
| 265 | or allows complexity and focuses on explaining just the out- | 319 |
| 266 | puts (<i>post hoc</i>). Scope indicates the range that a technique | 320 |
| 267 | possesses, i.e. if it explains individual predictions (<i>local</i>) | 321 |
| 268 | or the model’s entire behaviour (<i>global</i>). Finally, the ap- | 322 |
| 269 | plicability metric classifies techniques based on their model | 323 |
| | affinity, i.e. whether they are only compatible with a spe- | |
| | cific family of models (<i>model-specific</i>) or virtually any kind | |

324 form this step beforehand and not during the inference stage
325 (for speed reasons).
326

327 After obtaining the synthetic images and their respective
328 masks, the synthetic dataset is structured based on the iris
329 positions, enabling faster search. To that end, the clustering
330 algorithm K-Means is trained on a subset of the iris seg-
331 mentation masks to compute three centroids, one for each
332 major iris position (i.e. left, centre and right). With them,
333 one can store the images based on their combination of iris
334 positions (e.g. left-left, right-centre ...). By doing so, when
335 searching, the algorithm can just rely on the synthetic pairs
336 that share the same combination as the test pair, saving time
337 and useless calculations. Obviously, the centroids provide a
338 liberal classification of where an iris is. During the search
339 process, the irises are subject to a stricter comparison.
340

341 Upon settling for a portion of the synthetic dataset that
342 closely meets the iris position constraint, the segmentation
343 masks are further used to determine which generated pairs
344 have the iris and eyebrow in, approximately, the same position
345 as their counterparts in the test pair. Such pre-condition
346 is a key to obtain visually pleasant explanations, given that
347 pixel-wise differences are extremely sensitive to differences
348 in phase (i.e., component misalignment). Accordingly, the
349 synthetic neighbour's score is given by:
350

$$\text{score} = \omega_{\text{masks}} * \|\text{test_pair}_A - \text{neighbour}_X\|_2, \quad (1)$$

351 being $\|\cdot\|_2$ the $\ell - 2$ norm and ω , a weight that considers
352 component misalignment. This way, we obtain a weighted
353 distance for each synthetic neighbour, with respect to the
354 first image of the pair (i.e. image A for the test pair and
355 image X for the synthetic one). Moreover, ω_{masks} serves
356 to favour pairs that have good alignment, using a factor of
357 $1 - \text{IoU}(\cdot, \cdot)$, i.e., the complement of the intersection-over-
358 union of both segmentation masks. Then, if the score is
359 smaller than the currently saved pairs, the considered pair is
360 kept as one of the best matches, up to that point. This iterative
361 process continues until every considered synthetic pair
362 is analysed. In practice, the search process finds, amongst
363 the (large) thousands of synthetic pairs, the ones closest to
364 the test pair, in terms of the first image. Therefore, given
365 that the second image of the test pair is not a *genuine* match,
366 it will most likely be different in some areas to those of the
367 optimal synthetic neighbours, and that is exactly the kind of
368 dissimilarities that make up the final explanation.
369

370 From this point forward, the K closest neighbours are
371 sorted according to their element-wise distance to image B ,
372 using (1) (obtaining the distance with respect to the second
373 images, instead). Finally, to produce the final explanation,
374 the K best neighbours are used to obtain the pixel-wise differ-
375 ences against the test pair image B . In practice, a neighbour
376 distance is subtracted from the total sum of distances,
377 creating an inverted distance. This assures that neighbours
with smaller distances are more important to the final result

378 than those with bigger distances. Then, the inverted dis-
379 tances are divided by the sum of inverted distances so as
380 to normalise them. The final difference image yields from
381 such inverted distances, acting as weights to determine the
382 importance of each intermediate difference. At the end, to
383 add visual appeal, the resulting explanations are coated with
384 red and green tones.
385

3.3. Implementation Details

386 The DenseNet-161 model was trained for 15 epochs with
387 a learning rate of 0.0002 and a batch size of 64 image pairs.
388 The Adam optimiser was used for the weight optimisation
389 process (with default β_1 and β_2 values). A similar training
390 setup was used to train the ResNet-18 model, albeit for a
391 smaller number of epochs (i.e. 5). For the Mask R-CNN's
392 training process, we kept its default values, translating into
393 a learning rate of 0.001, a batch size of 1 and 30 epochs
394 worth of training (in this case, fine-tuning from the COCO
395 pre-trained weights). Regarding the StyleGAN2 architec-
396 ture, its training comprises a total of 80.000 iterations and a
397 batch size of 8. After converging, the generator is capable
398 of synthesising realistic looking images, such as the roughly
399 400.000 pairs that make up the artificial dataset. Finally, for
400 the number K , that determines how many synthetic neigh-
401 bours should be kept, we used a default value of 15.
402

4. Experiments and Discussion

4.1. Datasets

411 As mentioned above, the proposed framework consists
412 of two modules, one for recognition and the other for expla-
413 nation purposes. Regarding the former, the chosen CNN is
414 solely trained on the UBIPr dataset [16], which is naturally
415 oriented towards periocular recognition problems and con-
416 tains valuable ID information. As for the latter, it mainly
417 relies on a combination of UBIPr and FFHQ [21]. Despite
418 not being immediately applicable to the context of this work
419 (i.e. it contains full face images, thus requiring extra steps
420 to extract the periocular region), the FFHQ dataset contains
421 unquestionable variety in terms of attributes, some of which
422 are scarcer in the UBIPr dataset. In practice, a small, but cu-
423 rated, portion of the FFHQ samples is used to create a more
424 varied super set (Fig. 3). Regardless of their source, all im-
425 ages are resized to a common shape, depending on the task
426 (i.e. 512x512x3 for Mask R-CNN, 256x256x3 for Style-
427 GAN2 and 128x128x3 for the CNNs).
428



432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
Figure 3. Samples from the two datasets used. The top row represents images of the UBIPr dataset, whereas the bottom row illustrates cropped samples of the FFHQ dataset.

4.2. Working Scenario

As it is usual in biometric recognition systems, it is important to define the working mode and world setting, upon which the system is built. With respect to the working mode, a system is said to run in verification mode (also referred to as *one-to-one*) if it tries to validate a claimed identity (i.e. subjects identify themselves and the system's task is to validate those claims by comparing the extracted features to those stored in a database). On the other hand, if in identification mode (*one-to-many*), a system will try to retrieve the most likely match, within a known set of possible IDs [1]. As for the world setting, it discriminates between systems that can only accept a pre-defined set of IDs (closed-world) and those that continue to function even when new, unexpected subjects appear (open-world).

Based on the definitions above, our method performs verification (the comparisons are done *one-to-one*) within an open-world setting, meaning that unseen subjects do not impose compromises or limitations.

4.3. Qualitative Evaluation

As for most explainable techniques, the proposed method yields explanations that have some degree of subjectivity. Nonetheless, a DenseNet-121 model, trained to perform the verification task, is paired with either LIME, SHAP or Saliency Maps to create a comprehensive comparison scheme, to which we add the method described in [9].

Fig. 4 displays the expected results from a visually explainable system. In practice, LIME tries to keep the most important super-pixels, SHAP highlights those it deems important in red tones and Saliency Maps produce greyscale explanations. As for the method by Huang and Li, it generates a heat-map in which red tones elevate important areas. Focusing on the common pairs between all methods, the left sample is essentially different with regards to eyebrow thickness and presence/absence of a noticeable skin spot. As for the right one, the most obvious disparities have to do with the eyebrow areas. Overall, our results are the most informative, when compared with the remaining four solutions. While LIME and SHAP do a decent job, Saliency

Maps provide a faint explanation. It is Huang and Li's method that comes closer to our level of visual appeal, by clearly highlighting portions of the eyebrow and a portion of subject *A*'s skin spot, in the left pair. Moreover, when given the right sample, it generates a solid red area comprising subject *B*'s eyebrow. However, upon closer inspection, our results show more appealing visual cues: in the left sample, distinct red tones on top of *A*'s skin spot and eyelashes, as well as, reiterated eyebrow differences in the right sample with highlights in both eyebrows, rather than just one. As for the remaining samples, the third (just below the first) is clearly explained by highlighting the entirety of both skin areas, which are obviously different between images *A* and *B*. Finally, in the fourth pair it is also shown how the eyelids differ, by colouring that periocular component on subject *B*'s image, and, in the fifth sample, subjects *B*'s eyebrow and iris are accurately shown in red.

4.4. Quantitative Evaluation

A comparison between the proposed method and an existing periocular recognition alternative is summarised in Table 1. A bootstrapping methodology was employed by sampling 90% of the available dataset and dividing the resulting samples between two subsets: training (80%) and test (20%). Then, the CNN was trained as usual and its EER and AUC scores were saved. Such process was repeated 10 times, culminating in means and standard deviations for both metrics. Note that, for these experiments, our method was trained using the UBIRIS.v2 dataset [8], following the same scheme as the other method. Analysing the obtained results, one can conclude that the recognition module of our framework surpasses its competitor with regards to the EER metric, even when considering an open-world setting (which the other method does not). It should be noted that, due to a modular design, the recognition module in the proposed framework can be replaced to achieve superior performance, while maintaining the explainability properties intact.

| Method | EER | AUC |
|---------------------|------------------------------------|------------------------------------|
| Ours (open-world) | $0.108 \pm 3e-2$ | $0.813 \pm 5e-2$ |
| Ours (closed-world) | $0.087 \pm 2e-2$ | $0.910 \pm 2e-2$ |
| Zhao and Kumar [23] | $0.109 \pm 2e-3$ | — |

Table 1. Comparison between the proposed method (in both world settings) and a state-of-the-art strategy (strictly operating in a closed-world setting).

4.5. Ablation Studies

The two major hyper-parameters of the proposed method are the number of neighbours considered (K) and the length

486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539

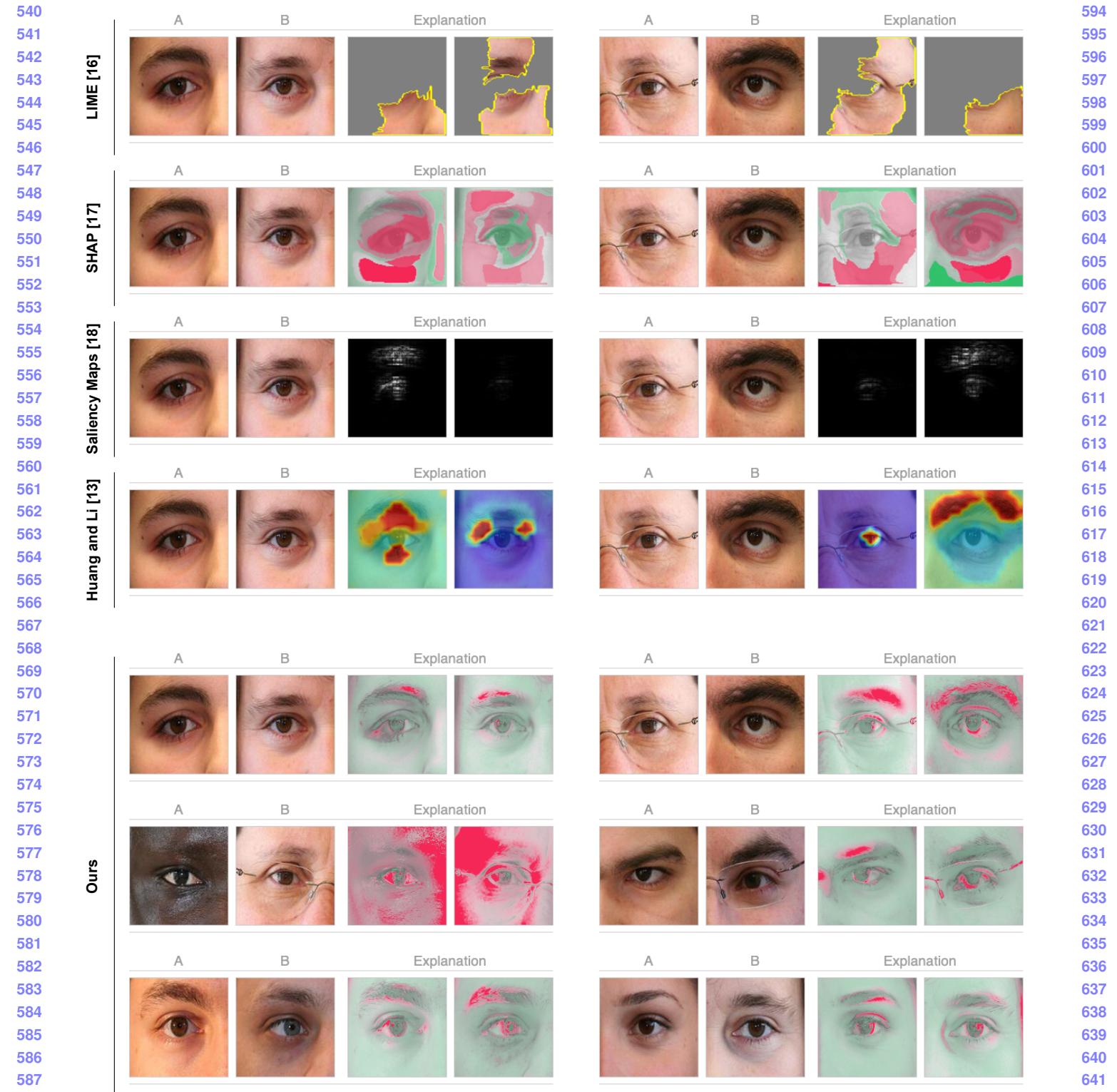


Figure 4. Comparison between the results attained by three standard interpretability techniques (LIME, SHAP and Saliency Maps), a state-of-the-art interpretable deep model for fine-grained visual recognition (i.e. [9]) and our method. Notice how our results are clearer in highlighting the components that justify every non-match decision (e.g., skin texture and color eyebrows/eyelashes size and distribution, irises color and even skin spots).

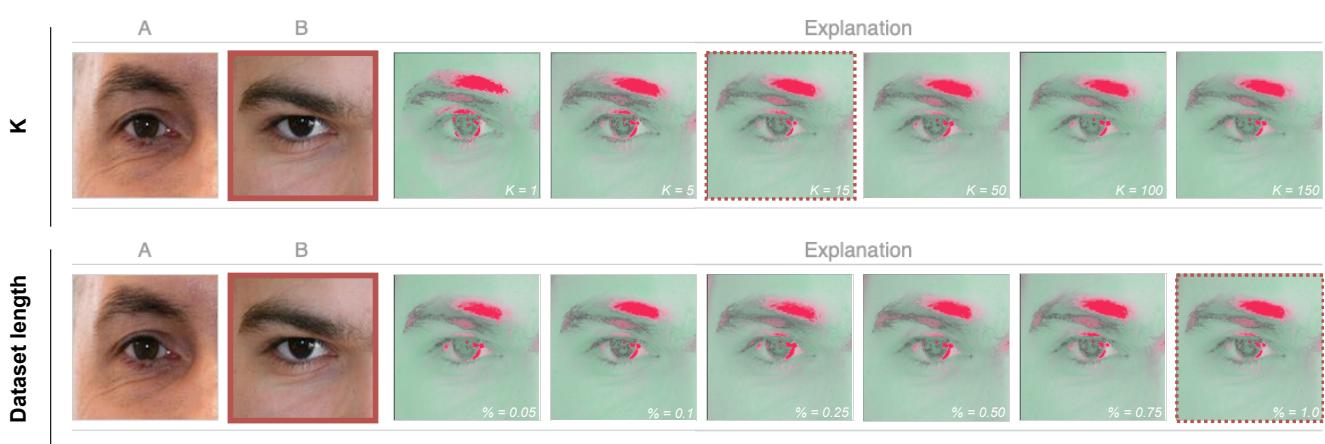


Figure 5. Visual perturbations when the two most important parameters of the proposed method are changed. The red square indicates which image is being *explained* (i.e. *B*), while the red dashed squares provide the default values used in our experiments. In general, increasing K up to 15 allows for smoother explanations, as does keeping a relatively large dataset. Reducing the latter tends to produce less sensitive results, substantially decreasing the plausibility of the visual explanations generated.

of the synthetic dataset. This section discusses how changes to these values affect the quality of the generated explanations in a less than optimal way (as seen in Fig. 5).

4.5.1 Number of Neighbours

The value K determines how many synthetic pairs should be considered the closest *genuine* neighbours to a test pair classified as being *impostor*. Overall, smaller values lead to more sensitive and jagged results, unlike those achievable with larger values. Up to a certain point (e.g. 15), increasing K creates smoother explanations, due to the larger number of samples taken into account when averaging the intermediate differences. This trend, however, starts returning incremental improvements (or none at all), that do not justify themselves (notice how, in Fig. 5, the output with K set to 50 or more stops presenting a prominent tone on the eyelid).

4.5.2 Length of the Synthetic Dataset

Considering the main goal of the search process is to find *genuine* pairs that closely resemble a test pair, restricting the amount of possible matches can impose certain limitations. With effect, in Fig. 5, it becomes clear how working with a smaller set of possible matches leads to less evident highlights, especially around the eyelid and, at a smaller scale, the eyebrow. Therefore, due to the use of relatively light calculations (i.e. MSE and IoU), the increased search times, that accompany larger datasets, do not make this an unfeasible solution, while also allowing for better results.

5. Conclusions and Further Work

This paper described an integrated framework - based in well known deep-learning architectures - to simultaneously perform periocular recognition and - most importantly - to provide visual explanations of the regions/features that sustained a particular *non-match* decision. By harnessing the generative power of GANs to create a very large set of synthetic pairs that come from the *genuine* distribution, for every *impostor* comparison we are able to perceive the regions and features that *failed the most* (i.e., those most evidently different from the a selected subset of the genuine synthetic set). This enables to generate clear visual explanations, where each component of the periocular region appears with a different colour depending of its similarity between a specific images pair. Importantly, the modular nature of the proposed method ensures that, the periocular region can be replaced by other biometric traits (e.g., the face) without compromising the explanations.

References

- [1] A. Ross A. K. Jain and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004. 1, 5
- [2] D. Braines R. Tomsett A. Preece, D. Harborne and S. Chakraborty. Stakeholders in explainable ai. arXiv:1810.00184 [cs], Set. 2018. Accessed on: Feb. 27, 2021. [Online]. Available: <http://arxiv.org/abs/1810.00184>. 1
- [3] D. W. Apley and J. Zhu. Visualizing the effects of predictor variables in black box supervised learning models. arXiv:1612.08468 [stat], Ago. 2019. Accessed on: Feb. 27,

- 756 2021. [Online]. Available: <http://arxiv.org/abs/1612.08468>.
757 3
- 758 [4] H. Li X. Shen B. Yin, L. Tran and X. Liu. Towards interpretable face recognition. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9347–9356, 759 2019. 3
- 760 [5] European Commission. General data protection regulation. 761 2018. Accessed on: Feb. 27, 2021. [Online]. Available:
762 <https://gdpr-info.eu>. 2
- 763 [6] A. Ross et al. Matching highly non-ideal ocular images: An 764 information fusion approach. *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 446–453, 2012. 3
- 765 [7] Jerome H. Friedman. Greedy function approximation: 766 A gradient boosting machine. *The Annals of Statistics*, 767 29(5):1189–1232, 2001. 3
- 768 [8] R. Santos J. Oliveira H. Proen  a, S. Filipe and L. A. Alexandre. The ubiris.v2: A database of visible wavelength iris 769 images captured on- the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 770 32(8):1529–1535, 2010. 5
- 771 [9] Z. Huang and Y. Li. Interpretable and accurate fine-grained 772 recognition via region grouping. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 773 pages 8659–8669, 2020. 3, 5, 6
- 774 [10] A. Vedaldi K. Simonyan and A. Zisserman. Deep inside 775 convolutional networks: Visualising image classification 776 models and saliency maps. *arXiv:1312.6034* [cs], Abr. 777 2014. Accessed on: Feb. 27, 2021. [Online]. Available:
778 <http://arxiv.org/abs/1312.6034>. 3
- 779 [11] A. Kumar L. Nie and S. Zhan. Periocular recognition using 780 unsupervised convolutional rbm feature learning. *2014 22nd International Conference on Pattern Recognition*, pages 781 399–404, 2014. 3
- 782 [12] Z. C. Lipton. The mythos of model interpretability. 783 *arXiv:1606.03490* [cs, stat], Mar. 2017. Accessed on:
784 Feb. 27, 2021. [Online]. Available:
785 <http://arxiv.org/abs/1606.03490>. 3
- 786 [13] S. Lundberg and S. Lee. A unified approach to interpreting 787 model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, page 788 4768–4777, Red Hook, NY, USA, 2017. Curran Associates Inc. Long Beach, California, USA. 3
- 789 [14] S. Singh M. T. Ribeiro and C. Guestrin. “why 790 should i trust you?”: Explaining the predictions of any classifier. *arXiv:1602.04938* [cs, stat], Ago. 2016. Accessed on: Feb. 27, 2021. [Online]. Available:
791 <http://arxiv.org/abs/1602.04938>. 3
- 792 [15] C. Molnar. Interpretable machine learning. a guide 793 for making black box models explainable. 2019. Accessed on: Feb. 27, 2021. [Online]. Available:
794 <https://christophm.github.io/interpretable-ml-book/>. 3
- 795 [16] C. Padole and H. Proen  a. Periocular recognition: Analysis 796 of performance degradation factors. *Proceedings of the Fifth IAPR/IEEE International Conference on Biometrics – ICB 2012*, 2012. New Delhi, India. 4
- 797 [17] H. Proen  a and J. C. Neves. Deep-prwis: Periocular 798 recognition without the iris and sclera using deep learning frame- 799 works. *IEEE Transactions on Information Forensics and Security*, 13(4):888–896, 2018. 3
- 800 [18] H. Su M. Bennamoun S. Minaee, A. Abdolrashidi and D. Zhang. Biometrics recognition using deep learning: A survey. *arXiv:1912.00271* [cs], Fev. 2021. Accessed on: Feb. 27, 2021. [Online]. Available:
801 <http://arxiv.org/abs/1912.00271>. 1
- 802 [19] B. C. Dhara R. K. Rout S. Umer, A. Sardar and H. M. Pandey. Person identification using fusion of iris and periocular deep features. *Neural Networks*, 122:407–419, 2020. 3
- 803 [20] B. Mittelstadt S. Wachter and L. Floridi. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2):76–99, 2017. 2
- 804 [21] S. Laine T. Karras and T. Aila. A style-based generator architecture for generative adversarial networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019. Long Beach, CA, USA. 4
- 805 [22] A. Ross U. Park and A. K. Jain. Periocular biometrics in 806 the visible spectrum: A feasibility study. *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, pages 1–6, 2009. 3
- 807 [23] Z. Zhao and A. Kumar. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Transactions on Information Forensics and Security*, 12(5):1017–1030, 2017. 3, 5
- 808 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863