X-Faces: The eXploit Is Out There

João Correia and Tiago Martins and Pedro Martins and Penousal Machado
CISUC, Department of Informatics Engineering,
University of Coimbra,
3030 Coimbra, Portugal
{jncor, tiagofm, pjmm, machado}@dei.uc.pt

Abstract
In the combinatorial form of creativity novel ideas are produced through unfamiliar combinations of familiar ideas. We explore this type of creativity in the scope of Data Augmentation applied to Face Detection. Typically, the creation of face detectors requires the construction of datasets of examples to train, test, and validate a classifier, which is a troublesome task. We propose a Data Augmentation technique to autonomously generate new frontal faces out of existing ones. The elementary parts of the faces are recombined using Evolutionary Computation and Computer Vision techniques. The key novel contributions include: (i) an approach capable of automatically creating face alternatives; (ii) the creation and usage of computational curators to automatically select individuals from the evolutionary process; and (iii) an experimentation with the interplay between Data Augmentation and serendipity. The system tends to create a wide variety of unexpected faces that exploit the vulnerabilities of face detectors. The overall results suggest that our approach is a viable Data Augmentation approach in the field of Face Detection.

Introduction
Face Detection (FD) systems are being thoroughly researched due to their wide range of applications, including entertainment services, social networks, search engines, and security systems (Yang, Kriegman, and Ahuja 2002; Zhang and Zhang 2010). Typically, such detectors employ classifiers that are created using example-based learning techniques. In this way, the dataset plays a key role not only for attaining competitive performances but also for assessing the strengths and shortcomings of the classifier. This means that the creation of adequate datasets for training, testing, and validation of the classifier becomes a crucial process.

The creation of a face classifier requires the construction of a dataset composed of positive and negative examples. A positive example consists of an image containing at least one face and the respective bounds; a negative example consists of an image that does not contain faces. Collecting both types of examples presents troublesome endeavours as the annotation of several images containing faces is a tiresome task, and gathering negative examples is far from being a straightforward task due to the inexistence of well-defined guidelines (Sung and Poggio 1998; Machado, Correia, and Romero 2012b). For these reasons, after gathering a number of examples, Data Augmentation (DA) techniques are usually used to expand the dataset (Chen et al. 2007).

We propose an approach that aims the expansion of the dataset of positive examples through the generation of new examples out of the existing ones. The idea is to recombine the elementary parts of frontal faces, i.e. mouths, noses, eyes and eyebrows, using Computer Vision (CV) techniques. A Genetic Algorithm (GA) is used to automatically recombine these parts and this way create new faces that are different from the original ones. To guide the evolutionary process we resort to an automatic fitness assignment scheme that employs a classifier (Machado, Correia, and Romero 2012a; Correia et al. 2013). The evolutionary engine is designed so the activation response of the classifier is minimised. As such, the evolutionary process tends to evolve exploits of the classifier, i.e. faces that are no longer considered as faces by it (see, e.g., figure 1). We have also implemented computational curators to automatically select examples that are evolved during the evolutionary process. We consider the results interesting, diversified, and sometimes peculiar.

Figure 1: Face evolved by the system that is not classified as so by the classifier.
Our work is motivated by the combinatorial form of creativity, where novel ideas are produced through unfamiliar combinations of familiar ideas (Boden 2009) and by the idea of recognising new invented artifacts even when we are not specifically searching for them (Pease et al. 2013; Schorlemmer et al. 2014). Thus, the main contribution presented herein is an approach capable of automatically generate positive examples out of existing ones. Other contributions include: (i) the combination of CV and Evolutionary Computation (EC) techniques to automatically generate faces; (ii) the analysis of the results evolved using different classifiers trained under different conditions; (iii) the usage of curators to select individuals from the evolutionary process; and (iv) an experimentation with the interplay between DA and serendipity.

The remainder of this paper is organised as follows. We begin by summarising the related work. We proceed to the thorough explanation of the proposed approach. Then, the experimental setup is described and the results are analysed. Finally, conclusions and directions for future work are presented.

Related Work

In this section we analyse contributions on the topic of creation of human faces using EC techniques. We separate the contributions into two groups: the creation of faces out of existing ones; and the creation of faces from scratch.

Johnston and Caldwell (1997) have developed a criminal sketch artist by implementing a GA to recombine regions of existing face images. With the objective of improving a FD system, Chen, Chen, and Gao (2004) have employed a self-adaptive GA to manipulate face images using image segmentation techniques and photometric transformations. The transformations applied in these two approaches can be destructive and for this reason invalid solutions may be created. Frowd, Hancock, and Carson (2004) have used a GA in combination with Principal Component Analysis (PCA) and eigen vectors to create an Interactive Evolutionary Computation (IEC) system aimed to evolve human faces. Limitations of this approach include its dependence on the user guidance and the small variety of results that it can generate.

The second group of contributions include approaches that use a general purpose evolutionary art tool to evolve faces from scratch. Examples of such approaches include the contributions of DiPaola and Gabora (2009) and Ventura (2010) where the similarity to a target image is used to assign fitness. Both approaches suffer from the limitations of IEC and the solutions tend to be too abstract or too similar to the target image. In previous work, we have also used a general purpose evolutionary art tool and an automatic fitness assignment scheme to evolve images from scratch, including figurative images such as human faces and ambiguous images (Machado, Correia, and Romero 2012a; Correia et al. 2013; Machado et al. 2015). This approach tends to evolve abstract imagery and for this reason it does not guarantee that, from a subjective point of view, the results resemble a face.

We have extracted some observations from our analysis of the shortcomings of the related work that guided the design of our approach: (i) it should be able to explore the search space in an automatic way, avoiding the user fatigue; (ii) it should guarantee the generation of valid faces; and (iii) it should promote the creation of faces considerably different from the ones contained in the dataset.

The Approach

The proposed approach is part of the Evolutionary Framework for Classifier assessmenT and ImproVemenT (EFFICIENT) (Machado, Correia, and Romero 2012b; Correia et al. 2013) and is composed of three different modules: an annotation tool, an evolutionary engine, and a classifier. Thus, the approach comprises the following steps: (i) the annotation of training examples by indicating the bounds of the faces and their parts; (ii) the training of a classifier with these examples; and (iii) the automatic evolution of new examples using the classifier to assign fitness. The three different modules are detailed in the following subsections.

Annotation tool

We develop a general-purpose image annotation tool (see figure 2). It allows the user to annotate objects present on images. One can annotate an object by positioning a sequence of points along its contour and by choosing the corresponding category. New categories can be added at any moment. The annotations created by the user are automatically saved in output files, more particularly in one eXtensible Markup Language (XML) file for each image and in one text file for each object category. The tool also exports the mask of each annotated object. When one opens a folder with images, the tool loads the corresponding annotations saved in files if they exist.

![Figure 2: Screenshot of the annotation tool. A demo video can be seen at http://cdv.dei.uc.pt/2016/x-faces-annotation-tool.mov](http://cdv.dei.uc.pt/2016/x-faces-annotation-tool.mov)
We use this tool to annotate the elementary parts of faces on a set of images. In this work, each face is annotated by indicating the bounds of its left eye, right eye, left eyebrow, right eyebrow, nose, mouth, as well as the bounds of the face itself.

Classifier

We train a cascade classifier to detect frontal faces based on the work of Viola and Jones (2001). It uses a set of small features in combination with a variant of Adaboost (Freund and Schapire 1995) to attain an efficient classifier, and assumes the form of a cascade of small and simple classifiers that use Multi-scale Block Local Binary Patterns (MB-LBP) features based on the work of Liao et al. (2007). Local Binary Patterns (LBP) features are intensity and scale invariant image descriptors. The original LBP features, introduced by Ojala, Pietikäinen, and Harwood (1996), label the pixels of an image by thresholding a rectangular region (e.g. $3 \times 3$) neighbourhood of each pixel with the centre value and considering the result as a binary string or a decimal number. MB-LBP extend the concept of LBP features to different sub-regions (blocks) around a center block, where the average intensity values of all the blocks are calculated. Then the LBP feature is extracted from the averages. The FD algorithm employs this classifier and can be summarised in following steps:

1. Define $w$ and $h$ as the width and height, respectively, of the input image.
2. Define a window of size $w' \times h'$, e.g. $20 \times 20$.
3. Define a scale factor $s$ greater than 1. For instance, a scale factor of 1.2 means that the window will be enlarged by 20%.
4. Calculate all windows with size $w' \times h'$ from the position $(0,0)$ to $(w-w',h-h')$ with 1 pixel increments of the upper left corner.
5. Apply the cascade classifier for each window. The cascade has a group of stage classifiers, as represented in figure 3. Each stage is composed of a group of MB-LBP features that are applied to the window. If the overall resulting value is lower than the stage threshold, the classifier considers that the window does not contain a face and for this reason terminates the search. If it is higher, it continues to the next stage. If all stages are passed, the window is classified as containing a face.
6. Apply $s$ to $w'$ and $h'$, and go to step 4 until $w'$ exceeds $w$ or $h'$ exceeds $h$.

Evolutionary Engine

The evolutionary engine is a conventional GA where the individuals are faces constructed from parts of different faces. Figure 4 explains the genotype of each individual and its phenotype. Each genotype is mapped into a phenotype by creating a composite face, i.e. the face parts encoded in the genotype are placed over a base face that is also encoded in the genotype. This process is accomplished by using a CV clone algorithm that allows the seamless placement of an image upon another (Pérez, Gangnet, and Blake 2003).

To guide evolution we adopt a fitness function that converts the binary output of the face classifier to an output that provides a suitable fitness landscape. This is attained by accessing internal values of the classification task that give an indication of the degree of certainty in the classification. In this work, we are interested in a fitness function that penalises individuals that are classified as faces. As such, the fitness function is defined as:

$$ f(x) = (tstg - pstg) + (tstg * ndet) + \frac{1}{1 + \text{stgdf}} $$

where $tstg$ is the total number of stages of the classifier, $pstg$ is the number of stages that the input image passes, $ndet$ is a boolean variable that tracks if no face is detected in the image, i.e., if the generated face is not detected by the classifier it yields a value of 1; and $\text{stgdf}$ is the difference between the value attained in the last stage that the image passes and the threshold of that stage.

Experimental Setup

Since we are interested in evolving faces from existing ones, in a similar way as in a GA approach, our objective is to evolve positive examples that are misclassified as negative examples, i.e. faces that are classified as not faces. We begin by defining two datasets of positive examples, one with 200 examples and the other with 500 examples. We then use these two datasets to train the classifiers face200 and face500, respectively. The face200 dataset includes all 200 annotated examples that are used in the GA for recombination. The face500 dataset contains all the face200 examples plus 300 other examples. With these two datasets we intend to explore the impact of our approach in a scenario where all
available faces have their parts annotated and in a scenario that only a fraction of the available instances are annotated. Furthermore, we are interested in the analysis of the resulting individuals in both scenarios.

The positive examples that compose both datasets were extracted from the FACITY project, which is a world wide project that gathers the pictures of photographers capturing the multiplicity of human faces from different cities and countries. We maintain the same negative examples used in previous experiments, composed of 1905 images from the ‘‘Urtho – Negative face dataset’ which contains different types of images including landscapes, objects, drawings, and computer generated images. In the scope of this paper, we intend to study the results of our approach while using face200 and face500 to assign fitness.

Table 1: Training Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example width</td>
<td>64</td>
</tr>
<tr>
<td>Example height</td>
<td>64</td>
</tr>
<tr>
<td>Number of stages</td>
<td>20</td>
</tr>
<tr>
<td>Min. hit rate per stage</td>
<td>0.999</td>
</tr>
<tr>
<td>Max. false alarm per stage</td>
<td>0.5</td>
</tr>
<tr>
<td>Adaboost algorithm</td>
<td>GentleAdaboost</td>
</tr>
</tbody>
</table>

We use the opencv traincascade tool of OpenCV to train each classifier. The main classifier parameters can be consulted in table 1 and were chosen based on the works of Viola and Jones (2001) and Lienhart, Kuranov, and Pisarevsky (2002). As for the FD settings we use the default parameters of OpenCV, which are presented in table 2. The test of each parameter is beyond the scope of this paper and besides that the default parameters enable a compromise between performance and speed of detection (Lienhart, Kuranov, and Pisarevsky 2002).

Table 2: Detection Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale factor</td>
<td>1.2</td>
</tr>
<tr>
<td>Min. face width</td>
<td>0.7 \times \text{example width}</td>
</tr>
<tr>
<td>Min. face height</td>
<td>0.7 \times \text{example height}</td>
</tr>
</tbody>
</table>

We test two experiments: exp200 and exp500. In exp200, face200 is used to guide evolution and face500 to curate individuals. In exp500, face500 is used to guide evolution and face200 to curate individuals. The role of the curator is to observe the evolved individuals and to select relevant faces that it does not classify as faces. We study the behaviour of each curator and its selection of individuals.

The evolutionary engine settings are presented in table 3. In terms of face parts recombination, we maintain the pairs of eyes and the pairs of eyebrows, reducing the genotype length from seven to five. The rationale for this decision is related with the fact that most faces have a certain horizontal symmetry that the classifier tends to learn from the positive examples. In preliminary experiments, we have observed that the classifiers struggled on images where the pair of eyes and eyebrows belonged to different faces, leading to an early convergence of the evolutionary process. Besides the technical aspects, the images evolved were unnatural and easily noticeable that were blends.

Experimental Results

In this section we present and analyse the experimental results. We begin by analysing the evolution of fitness in the two experiments. Afterwards, we discuss the progression of detections over the generations. Then, we present and discuss the individuals selected by the curators. Finally, we analyse the visuals of the evolved individuals.

Figure 5 shows the evolution of fitness of the best individuals along the generations in exp200 and exp500. For each experiment, we plot both fitness curves to examine how one affects the other. We can observe that the evolutionary algorithm is able to optimise the fitness function. In both experiments, when the fitness value of the guiding classifier increases, the fitness value of the curating classifier tends to behave similarly. The values reveal that it is easier to satisfy face200 either when it is evaluating or curating. The observed behaviour in the exp500 plot suggests that even being the face500 that is guiding, the fitness values of face200 are higher than the ones of face500. On the other hand when face200 is guiding, the fitness that uses face500 also increases, suggesting that exp200 is also suitable of evolving solutions for the exp500. As we are interested in evolving individuals that are not classified as faces, one can say that we are promoting the evolution of examples that are not present in the training datasets. Bearing this in mind, while reflecting on the training datasets of both experiments, the results suggest that the evolved individuals in exp500 tend to be new in the perspective of face200, i.e. different to the ones present in the face200 training dataset. The behaviour of face500 in exp200 suggests that it is able to evolve individuals that are new in the face500 dataset but at a smaller rate. This can be a consequence of the face200 training instances being included in the face500 training.

In figure 6 we observe the average of individuals that are classified as faces throughout generations. The number of detections decreases in both experiments, showing the ability of the approach to evolve faces that are no longer classified as such. In both experiments, face500 obtains higher
Figure 5: Evolution of the fitness of the best individual throughout generations when using face200 to guide evolution and face500 to curate individuals (top); and the other way around (bottom). The visualised results are averages of 30 runs.

number of detections than face200. This difference is more pronounced in exp200. As for the progression of the curves, one can see that in exp200, face200 decreases at a higher rate. In exp500, both curves decrease at a similar rate. One can also say that in exp200 the evolution promotes the generation of solutions that are classified as faces by face500. We consider that this is consistent with the fitness curve behaviour of figure 5 suggesting that the system exploits vulnerabilities common to both classifiers.

Figure 6: Progression of the average of detections when using face200 to guide evolution and face500 to curate individuals (top); and the other way around (bottom). The visualised results are averages of 30 runs.

curated individuals in exp200 (see figures 8 and 9). When face500 is curating or guiding, the evolved individuals share more characteristics (see figures 9 and 10). This is consistent with the idea that the exploits of face500 tend to be also exploits of face200.

Figure 7 depicts in which generation the fittest individual, on average, seize to be classified as face. One can conclude that when face200 is guiding, it takes less than 10 generations for the best individual to stop being classified as a face. The behaviour of face500 suggests that the fittest individuals are still classified as faces in the final generations. In contrast, when face500 is guiding, the results indicate that there are evolutionary runs where the fittest individual is still classified as a face by both classifiers.

Figure 7 depicts in which generation the fittest individual, on average, seize to be classified as face. One can conclude that when face200 is guiding, it takes less than 10 generations for the best individual to stop being classified as a face. The behaviour of face500 suggests that the fittest individuals are still classified as faces in the final generations. In contrast, when face500 is guiding, the results indicate that there are evolutionary runs where the fittest individual is still classified as a face by both classifiers.

Figures 8 and 10 depict a selection of fittest individuals registered in the experiments. As for figures 9 and 11 one can observe some of the curated individuals. The results suggest that although there are individuals in common, the two curators tend to select different individuals. Some of the selected faces share characteristics that we consider as exploits of the classifiers, particularly at the level of the skin tone, contrasts, and size of some facial features. One can conclude that there are overlaps between the fittest and the curated individuals in exp200 (see figures 8 and 9). When face500 is curating or guiding, the evolved individuals share more characteristics (see figures 9 and 10). This is consistent with the idea that the exploits of face500 tend to be also exploits of face200.

Figure 8 depicts a selection of individuals evolved in different runs that we found to be interesting and peculiar. This selection shows the ability of the approach to explore the search space and exploit the vulnerabilities of the classifiers in an automatic and tractable way. As such, one could expect simple recombinations of faces that the classifier has not "seen" before or exploits of lighting and contrast conditions. Nevertheless, the system produces atypical faces with unexpected features. For instance, one can see convincing images of babies with piercings, cases of gender ambiguity, and mixtures of interracial attributes that are at least visually uncommon and peculiar. Some of the generated faces are so realistic but disturbing at the same time that one could relate with the uncanny valley problem MacDorman et al. (2009), i.e., the phenomenon where computer generated figures or virtual humanoids that approach photo-realistic perfection make real humans uncomfortable.

Based on the notion of serendipity by Pease et al. (2013), the system with a prepared purpose, based on previous knowledge obtains a new result that is suitable and useful for the system and external sources. The purpose of this DA approach is to generate new useful examples of faces, in this case, faces that are not detected by the classifier that
is guiding the evolutionary process. Most of the outputs of this process are evaluated by the guiding classifier, by the computational curator, and by ourselves as new, valuable, interesting and unexpected. Since the system starts with a set of limited pre-defined parts, some individuals evolved suggest the occurrence of serendipity. Based on the ability of the system to generate such examples, one could insert domain knowledge to identify and enforce the generation of these examples. One could also use a curator or multiple curators to take an active part in the generation process, i.e. a component in the fitness function.

Although our approach generates faces that are not detected by the classifiers, the images can be used in other domain of applications. They can be used as inspiration to video games or movie making in the form of characters, and in visual arts to generate crowds with different faces. One could easily adapt the system and consider other items into play. For instance, the placement of the face parts could be modified for a horror scenario, allowing the face parts to be placed in different positions, e.g. replace all face parts with eyes, replace the two eyes with two mouths, and exchange the eyebrows with mouths. Due to the primary goal of this work, we only use human faces and their parts. However it is possible to use parts from other contexts, e.g. add and recombine parts from other animals and objects, allowing the creation of surrealistic artifacts. Our system can use multiple object classifiers to analyse photos and manipulate, mismatch, place, and edit the detected objects or introduce new ones.

A final comment goes for the potential use of this approach for DA. Similar to typical bootstrapping, this approach can generate variations or completely new examples from a pre-defined sub-set. The generated examples may be used to further improve the quality of the training dataset and thus the quality of the classifier, a path that is already being pursued.

**Conclusions and Future Work**

We have described and tested an approach for the automatic generation of faces based on the principles of combinatorial form of creativity. The experimental results demonstrate the ability of this approach to generate a wide variety of faces that test the ability of the classifiers to detect them. As such, we consider the approach proposed herein a viable solution for DA in the field of FD. The results also show the impact
of different classifiers on the evolved faces. Besides fulfilling its purpose, from our perspective, the faces created have interesting and unexpected features.

The proposed approach may benefit from future enhancements, including: (i) the implementation of automatic detection and landmark mechanisms in the annotation tool to assist the annotation of the face parts; (ii) the use of the evolved faces to extract more face parts; (iii) the further integration of the proposed approach with EPECTIVE, so the evolved examples are used to (re)train the classifiers in an attempt to improve their performance; (iv) the exploration of different curators; and (v) the expansion of the approach to other problems or scopes.

**Acknowledgement**

This research is partially funded by Fundação para a Ciência e Tecnologia (FCT), Portugal, under the grants...
References


