### **Improving Forensic Face Aging to Aid Search for Lost and Missing Children**

Fall 2023 DS549 Machine Learning

#### **Project Description**

[Atfal Mafkoda](https://atfalmafkoda.com/en/home) (translates to *Missing Children)* is a community project based in Egypt interested in applying age progression forensics to pictures of the missing children, since some of them have been missing for years. The project has had impressive success stories in finding missing children (and adults) whether gone missing, abducted and used in human trafficking and child abuse [*see Additional information*]. The project primarily relies on its Facebook page where the administrators post the missing person’s photo and most of their known features and the community would message back if any member could identify the whereabouts of the person. The working volunteers would visit the place and attempt to confirm that this is actually the same person and in such case, confirm with the family any extra details and features and help return the person to their family. While the process may seem too simple to be true, it has successfully returned 3000+ missing persons to date.

However, the process is not effective when the photos are too old, picture quality is low, and the children have gone missing for too many years that their photos no longer match their current facial features. In those circumstances and in the case of abduction, there is little chance that the children would be found.

*Note: Unfortunately, in developing countries many of those missing children may end up living on the streets. A final filter that would apply the effects of living on the streets such (dirty skin, skin imperfection, jagged unshowered hair) would also help the community in identifying those children. In the US, it may be able to help identify runaway teenages. Ease of utilization for the project would be a plus as it would then be integrated and available to the volunteers faster. Atfal Mafkooda would share any technology that is based on the output of this project that would further simplify utilization.*

***Project Context and Organization Description:***

[*Atfal Mafkoda Website*](https://atfalmafkoda.com/en/home) *For more information consult the Additional Information section.*

**Project Goal**

To address these challenges we propose integration into software we're building to automate finding missing children based on: 1. Images for missing kids we have in our DB and 2. Images for found Jone/Jane Doe's that we have in our DB

Tom and Oindrilla’s Recommendations:

1. Address data label limitations – the missing children’s photos are posted on Facebook with free form text in Arabic accompanying. Extract child names from the text to label the photos.
2. Continue the face identification model implementation with the labeled photos and incorporate aged faces as references.

***What we have now:***

* *Face rotation - to be front facing*
* *Aging images*
* *Image comparison*

***What we'd like to add:***

* *Family feature morphing*
* *Homeless feature morphing*
* *Optimization to run on low cost servers (explore quantization techniques to reduce cost)*

Continuation of applying generative deep learning methods to:

1. **Improve forensic face aging** to provide **higher quality aged face photos**. Note that some of the challenges here are the quality of the original photo, and the young age of the subjects. Existing face aging GANs may not target such young ages as starting points.
   1. Explore style and face feature transfer techniques to incorporate facial features of family members in the hopes of generating more realistic aged photos.
2. **Explore techniques to apply visual effects of possible homelessness to the aged photos** to account for the fact that the children may have been subject to the physical stresses of homelessness.
3. **Develop face matching/recognition techniques to narrow down large corpus of missing children photos to a small number of most likely matches**, in order to direct the attention of health workers and volunteers to the most likely matches.

#### **Project Steps**

Key Initial Steps

1. Review Project Description and determine questions or clarifications needed from the client.
   * 1. Read at least 3 of the sources in the Additional Information section
     2. Review previous team;s insights below:
     3. Previous Team’s Insights

* Before continuing with this project, it will be helpful to understand the pipeline and reproduce the previous results by having the pipeline running with appropriate environment configuration
* Keep in touch with the client and communicate about the expected deliverables. It will be helpful to prepare demos for the phase progress or presentations in the client meetings, and focus on high-level ideas rather than technical details
* Follow up with the client on the usage condition of the pipeline and try to improve the pipeline/add new features due to their feedback

1. Create a work plan with your team and determine the breakdown of tasks into weekly sprints on the Trello board (to be shared by Spark! Staff).

Technical Information

* Language: Python
* Public datasets that may be of use:
  1. <https://github.com/HasnainRaz/Fast-AgingGAN>
  2. <https://github.com/ravindrabharathi/FaceAgingGAN>
  3. <https://github.com/Gokkulnath/FaceAgingGAN>
  4. <https://github.com/AbuAbdULLAH-MuhammadAli/FaceAgingStyleGANs>
  5. Last semester output repository: <https://github.com/BU-Spark/ml-atfal-mafkoda-missing-children>
  6. Data collection/labeling site code (We can provide viewer access to students for the production site): <https://github.com/yousseb/atfal-site>
  7. AI code based on last semester output: <https://github.com/yousseb/atfal-ai>

In addition, we might have access to images from the non-profit itself, which may include:

1. Photos of currently missing children.
2. Photos of recovered children and their original missing child photo
3. Photos of family members from both #1 and #2.

#### **Dataset License Terms and Protection Requirements**

#### **Dataset Description**

#### **Expected Final Deliverable**

The team to create:

1. Ideally: we should be able to predict how people will look like if (homeless, long hair, short hair, with mustache ... etc) + Reliable Face Aging
   1. End face aging pipeline output is given to clients as a python class preferably using CPU only, but Jupyter notebooks that use GPU is ok as well.

The team should have determined some of the following approaches:

1. Improved forensic face aging GAN
2. Transferring facial features of family members to face aged photos.
3. Simulating the visual effects of homelessness on face aged photos.
4. Face matching/recognition algorithm to narrow the list of possible matches to a candidate.

#### **Additional Information**

1. Spring 2023 Team’s Takeaway:
   1. Next Steps
      1. Fine-tune the Face Aging GAN by experimenting with different hyperparameters.
      2. Collect more pictures and information of missing children. Encourage public participation on uploading more pictures. Build an online platform for picture upload.
      3. Design and build an online database for secure and stable picture storage.
      4. Build on the existing pipeline according to the client’s suggestions.
2. Articles ( [Family Assisted Young Children Face Aging](https://docs.google.com/document/d/1D0kx_U39Tqcpwm-Nmp9Rw_8-HtcY52O7p9akTJFB2nI/edit#) )
   1. [Egypt Facebook page raises hopes for missing children - BBC News](https://www.bbc.com/news/world-middle-east-53564935)
   2. [Missing children in U.S. nearly always make it home alive | Reuters](https://www.reuters.com/article/us-usa-missing-children/missing-children-in-u-s-nearly-always-make-it-home-alive-idUSBRE83P14020120426)
   3. [Facts & Stats - Child Find of America](https://childfindofamerica.org/resources/facts-and-stats-missing-children/)
   4. [https://www.insider.com/man-thought-mom-died-reunited-44-years-later-facebook-page-2022-12](https://www.insider.com/man-thought-mom-died-reunited-44-years-later-facebook-page-2022-12?fbclid=IwAR0aGF3jHbc7zkusmGX-PgXpDbTP3Dl-AsTsICxBQRjMb4fsfCil6zVz_6g)
   5. <https://www.bbc.com/news/world-middle-east-53564935?fbclid=IwAR0gdikdahzk7Y0QPQejTW2CpAAFTfP41nGgYDqSHodo2IQnuBLcWqNHtpw>
   6. [The Facebook page searching for Egypt's missing children](https://restofworld.org/2021/facebook-page-searching-for-egypt-missing-children/)
   7. [Egyptian media owner detained after trafficking and sexual assault claims | Global development | The Guardian](https://www.theguardian.com/global-development/2022/jan/13/egyptian-media-owner-detained-after-trafficking-and-sexual-assault-claims)
   8. <https://english.alarabiya.net/features/2018/02/13/Website-selling-children-based-on-hair-eye-color-sends-shockwaves-across-Egypt>
   9. [The Facebook page helping to find Egypt’s lost children | Arab News](https://www.arabnews.com/node/1382181/middle-east)

#### **Contact Information**

| **Role** | **First Name** | **Last Name** | **Email** |
| --- | --- | --- | --- |
| Client - Founder | Rami | Elgebali | [relgebali@hotmail.com](mailto:relgebali@hotmail.com) |
| Client | Youssef | Abukwaik | [youssef.adnan@gmail.com](mailto:youssef.adnan@gmail.com) [youssef.abukwaik@intel.com](mailto:youssef.abukwaik@intel.com) |
| Spark! Support | Michelle | Voong | [mvoong@bu.edu](mailto:mvoong@bu.edu) |
| Class Instructor | Thomas | Gardos | [tgardos@bu.edu](mailto:tgardos@bu.edu) |
| Class Instructor | Oindrilla | Chatterjee | oc@bu.edu |
| PM | Rithvik | Doshi | rithvik@bu.edu |
| TE |  |  |  |
| Teammate |  |  |  |
| Teammate |  |  |  |
| Teammate |  |  |  |
| Teammate + Team Rep |  |  |  |

### 

### **Fall 2023**

#### **Meeting Notes**

Meeting 11: Final Meeting

* Presentation of Final Results
* Total age graph
* Flow Chart of Architecture
  + Data flow from JSON to csv table of matches

Meeting 10: Student Meeting

* Need to work on validation?
* As such, need to do some experiments:
  + Using the missing folder, multiple images per case id
  + Ran the matching alg btwn missing folder and missing folder
  + For each image trying to match to the person to the remaining matches that aren’t in the pool
  + Images don’t seem to be that good, even when enhanced the images becomes very smoothened
* Aging-filter
  + Making headway on that, using same approach as last semester’s
* Documentation for Youssef
  + Feel free to reach out to PMs for any help

Meeting 9: Client Meeting

* Using through deepface
* OpenCV face detection model couldn’t detect faces properly
* MTCNN model for face detection
* (mtcnn (detection) + facenet (matching)) Changing models reduced false positives, 0.7 is the threshold
  + Threshold picked randomly
* Images: 943 missing, 500 unknown
* Images with no matches: problem with the images themselves or the models?
  + Possibly anomalies in images (ex. photos of photos on computer screen)
  + Edge case that might be handled by community or contacting families for better images, focus on the happy path
* Worked on prediction visualizer - to see the results (easier viewing than the csv)
* Final deliverable: jupyter notebook with documentation, tricks and notes, indicate final notebook that Youssef should be using

Meeting 8: Student Meeting

* Try to transform photos to get them detected?
* Another model actually started working well
  + [MTCNN](https://github.com/ipazc/mtcnn) started working
  + Will go forward with it, trying it in a local env rather than just Colab
* Tuning hyperparameters done
* Deploy requirements not given, so will go ahead with a frontend deployment
  + Streamlit vs Gradio
* User story: give a photo of an unknown individual, the model matches the photo to missing individuals (all possible matches)
  + Could try to trim the matches using a threshold, just confirm with Youssef
* Working on replicating last sem’s environment, but not working out
  + Working on local, getting import errors
* Age progression with identity-preserving conditional GANs
  + Working through it, updates next meeting

Meeting 7: Client Meeting

* Clarifying data labels
* Matching output
* Deepface results shown
* Need to reduce false positive rates
* Add aging filters on top
* Client side:
  + Not too many updates

Meeting 6: Student meeting

* JSON data extracted into the 5 image classes
* Concentrating John Doe and Missing + trying to match them
* Enhanced photos may be cropped so no need to rebox on those pictures after enhancing
* Extracted images of each case
  + John Doe assumption: missing
* Created a face matching algorithm
  + Returns pool of matches (variable length) with similarity score (VGG-Face-euclidean)
  + Deep Face isn’t detecting certain images
  + Enforce\_detection set to false
  + 20% of images in Missing dir are not detected
  + 50% in unknown not detected
  + Going to add homelessness feature
    - Not sure about how far we can go with it
    - In the research phase about that
  + Family morphing feature
    - Incorporate family features into faces
    - Ask for that database
  + Age progression
    - Previous year’s + another identity-preserving method

Meeting 5: Client Meeting

* Youssef has given the team JSON data
* Reorganized to feed it into the face recognition model
* Had trouble creating accurate bounding boxes
* Getting back similarity scores
* Worked on EDA process, trying to underline the missing values and translate arabic text
* Trying to improve the deepface model
  + Small drawbacks:
    - Multiple images given to the model, mostly correct but with a few mismatches
* https://github.com/BU-Spark/ml-atfal-mafkoda-missing-children/tree/face\_recognition/fall23/deepface
* Questions
  + JSON data uses lots of images and unable to upload it all to SCC
  + Only takes 2GB at a time, JSON file takes a lot of time to upload
    - Try using batch job (qsub)

Meeting 4: Student Meeting

* [Midterm Presentation](https://docs.google.com/presentation/d/1B2hkZ9DjM7BzQ1Pufjtmw7kqIR4k0gsR/edit?usp=share_link&ouid=107597783919889816671&rtpof=true&sd=true)
  + Went well
  + How to use the age cam? Two papers outlined in the research document
  + Waiting for feedback
* Questions about the data
  + Image data has been classified into different types
  + Deceased, John Doe, Missing, Reunited, Unclassified
    - Confused about unclassified, John Doe?
* Mahaveer + Varshith: JSON data to csv and classes
* Face recognition models using GPUs, but need to be able to use CPUs
  + May not need to train it as there are standard weights + models out there
* Gitika + Navya worked on the presentation
* Continued work on the images, EDA, continue tomorrow

Meeting 3:

* Met with Youssef
* Prioritizing Face Detection for Sunday deliverable

Student Meeting 1:

Questions for Clients:

* What kinds of optimization need to be performed? Depending on which the pipeline may have to change to accommodate optimization
* This is at the end of the deliverable pipeline, but looking to plan ahead
* Kabilan’s Insight: Optimize Model and Pipeline to be run on CPU because the servers don’t seem to be particularly powerful
  + Atfal Mafkoda are using Oracle Cloud Free Tier
  + Consider Bayes’ model that can run on a CPU versus a larger model
* Take a look through atfal-ai GH repo
* Starting with deploying client solution, how well that pipeline performs

Meeting: Kickoff

Date: Oct 6 2023

Attendance: All students, Youssef, Tom

Discussion Points:

* Introduction to the project
* Starting on Facial Recognition, DB + adding ‘homelessness’ facial features

Action items:



---------

**Summer 2023 Current Status of Work from Youssef**

***Currently:***

* We realized that our data requires organization in order to be properly utilized by AI applications. We started a full Facebook data-mining project for our page. This is going underway but we have sufficient data at this point to start the AI development part. Around 1000 cases have been categorized (missing/john doe/reunited/deceased) and linked to their corresponding photos. We used commercial scrappers from apify in order to bypass the Facebook banning.

***Pipeline Implementation from Spring 2023:***

Created this repository <https://github.com/yousseb/atfal-ai/> which I will be maintaining until I get the full pipeline working:

* Goal: The target at this point is to run on Oracle Cloud Free Tier offer - deployment steps are in the repo's wiki. Oracle is offering four ARM based servers for free and I think that I could squeeze our pipeline in there. I replaced some of the models with ones that would work faster on CPU and so far - OpenVino and onnx based models. Output seems to be OK so far. :)
* Instructions on how to run locally are on the github repo, but most importantly, I have a quick test site up (that I will take down later once we actually start using the endpoints)

Testing:

* Test site: <https://ai.atfalmafkoda.net/docs>
  + (open the endpoint, click "Try it out", input image\_url and request body when needed)
* Use the following image for testing (or your one of the Facebook page if you wish): <https://www.ppic.org/wp-content/uploads/Crowd-of-Diverse-People_800x528-768x512.jpg>
* It's a pipeline, so you need to do the calls in order:
  + 1. Face detection (need just image url) -> produces list of face rectangle
  + 2. Enhancer (needs same image\_url and any rectangle generated from step #1)
  + 3. Frontalizer (need special image since you need to serve enhanced image - use this url for now: <https://raw.githubusercontent.com/HRLTY/TP-GAN/master/data-example/001_01_01_041_07_cropped.png>)

API

* The API should be fairly responsive, but I doubt that it could handle multiple users hitting it at the same time.
* If it's slow, there's a chance that it's being simultaneously used - which is not the intended behavior as the use-case is that it will be doing a single inference at a time. So if it feels too slow, try it again later.

Additional Links:

10/6/23:

* <https://github.com/yousseb/atfal-ai/tree/main/shelf>
* <https://github.com/BU-Spark/ml-atfal-mafkoda-missing-children>
* <https://github.com/TencentARC/GFPGAN>

Oct 16, 2023 Client meeting

1. Construct database - check access to data
2. Morphing to tackle “homeless” category+
3. Focus on facial recognition, can use open-source datasets
   1. Pair-wise matching or match set A to set B