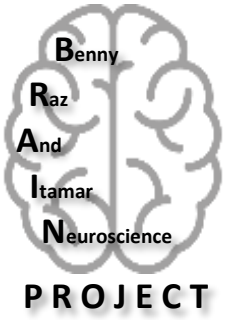


Grad-CAM visual explanation for face recognition

User manual

Benny Brazawoski, Itamar Kaminski, Raz Tamir

Table of Contents

Introduction ..... - 3 -

System overview ..... - 4 -

Minimal requirements ..... - 5 -

Web application sections and menus ..... - 5 -

Using the system - user workflow and system behavior ..... - 7 -

Troubleshooting ..... - 9 -

Introduction

While face recognition algorithms are at the state of the art level, with some performing better than human level in certain domains, it is not yet fully understood how they achieve their relatively high accuracy.

A main field of interest is in uncovering the underlying facial features that give rise to an algorithm prediction, as the features used by human in facial recognition are also unknown.

The best algorithms to date use machine-learning techniques consisting of deep artificial neural networks ("Deep Learning"). These networks are highly complex and are not human interpretable, consisting of millions of parameters being adjusted during the training phase of the algorithm.

As the main mission of neural networks is to make predictions on unseen samples, making those networks interpretable by it users is essential for two main reasons:

1. "Justification" – on correct predictions it helps us gain trust in the system (knowing the prediction was not accidental) and let it be part of automated processes.
2. Algorithm enhancing – on incorrect predictions it helps us gain insight on why the failure occurred, showing that seemingly unreasonable predictions have reasonable explanations, and therefore helping improving the algorithm itself and learn its fallbacks.

The interest in interpreting neural networks is not unique to face recognition; it is relevant to any field that uses deep learning algorithms. Yet each field need its own tailored way of justification. In the world of computer vision, the most intuitive way is highlighting the regions on the sample image that provide a "visual explanation" for a certain prediction.

The BRAIN project system provides visual explanation for face recognition algorithms, highlighting the facial features that gave rise for a certain recognition, namely giving an identity to a persons' face.

System overview

The BRAIN project system is a web-based application providing an easy to use application for feature visualization, providing visual explanation for face identity recognition.

Querying the system with a facial image will result with the following:

1. A name of a celebrity identity that is the most similar to the person in the query image.
2. Visual explanation - the query image with a layer of the visualization, highlighting the features that were most significant for the identification.
3. Verbal justification – a short textual explanation using facial landmarks.

Example:

Query image: uploaded by the user and sent to the server

Result: returned by the server

"You are most similar to ***Moshik Afiya*** *because of your* ***eyebrows****."*

The backend of the system is powered by deep learning algorithms, including state of the art networks for face identification and recognition, and a variation of a novel visualization algorithm called GRAD-CAM (which is originally made for object recognition algorithms).

Minimal requirements

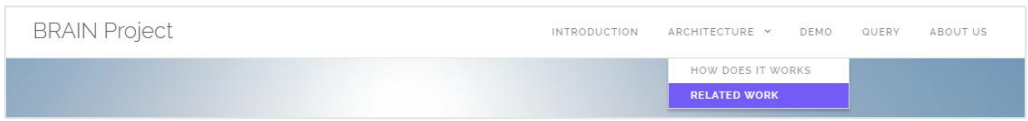
1. Internet connection
2. Web browser

Web application sections and menus

The web application consists with the following sections:

- Navigation bar
- Introduction
- System architecture and related work
- Demo
- Querying the system
- Results section
- About us

1. Navigation bar
    a. Provides a way to directly navigate between the different web application sections.
    b. Appearance
        i. In desktop\laptop computers – top menu bar
        ii. In a mobile device – drop-down menu



2. Introduction
    a. A short introduction in to the need for visualization and justification in facial recognition, and the main goals achieved by this project.

3. System architecture and related work

    a. A quick overview of the system backend architecture followed by links to resources of related work, offering the user to deepen his or her knowledge base with the topic.

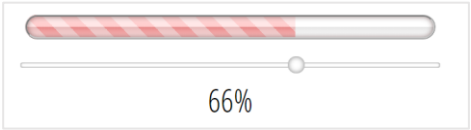
4. Demo section

    a. This section provides an annotated walkthrough demonstration of the system operation.

5. Querying the system

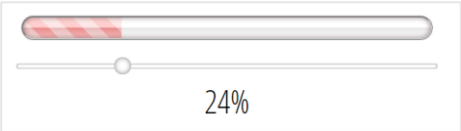
    a. This section is the heart of the application. Once the user is familiarized with the system, he is encouraged to start and play with it by uploading his own image and witness what the system result is.
    b. An upload field is used for the user to upload an image stored in the users' device, or if using a mobile device can capture an image via the devices' built-in camera.

    c. Adjustable parameters:
        i. Certainty level – an adjustable bar with values ranging between 0%-100%. Used by the user to indicate what the certainty level the system should use for recognition.



        ii. Number of results – a dropdown menu ranging between the values 1-3. Used by the user do define how many different identities the system should return.

iii. Visualization threshold - an adjustable bar with values ranging between 0%-100%. Used by the user to determine the threshold for the visualization layer.



6. Results section

   a. A section where the query results will be shown (appears only after a query is made).
   b. In this section, the user can see the number of results he asked for, with visual and verbal justification attached to each one of them.
   c. Social networks buttons are available for immediate sharing the results if desired.

7. About us

   a. This section provide some info about the authors of the project, accompanied with contact information.

Using the system - user workflow and system behavior

1. Entering the website is done by navigating to the following address:
   TBD

2. Once inside the website, navigation between sections can be done by clicking the navigation bar or just by scrolling down the page.

3. Introduction and architecture
    a. These first two sections are for those who are interested in a quick overview of the system.
    b. Those who are only interested to use the application may click the "demo" button in the navigation bar, or just scroll down.

4. Demo
    a. Three demo query images are shown, accompanied with walkthrough explanations.
    b. Clicking each one of the images will send a request to the server.
    c. The result returned by the server is a set of three images layered with GRAD-CAM visualization, each with a different pre-defined threshold for the GRAD-CAM algorithm.
    d. This demonstration is in order for the user to be acquainted with the GRAD-CAM visualization. The threshold can be adjusted later on the users' queries.

5. Querying the system
    a. Proceeding to the query section, which is the heart of the system.
    b. Once familiarized with the visualization system, the user can upload a query image (or take a live photo, if using a mobile device with a built-in camera).
    c. Adjusting the system parameters:
        i. Choose the number identities you wish for the system to return (the maximum is three).
        ii. Set the certainty level in which the system will only return identities that are above it.
        Caution: setting the certainty level too high might result in zero hits by the system.
        iii. Adjust the GRAD-CAM visualization threshold.
        Caution: setting it to its extreme values may result in meaningless visualization.

d. Press submit and wait for the results.
e. The server will return a result (as the number of identities requested by the user):
    i. The name of the identity with it certainty level.
    ii. The query image with the GRAD-CAM visualization layered on it.
    iii. The facial landmarks that signifies the highest explanation for the identity classification.
f. The result can be easily shared to any social network by clicking the appropriate button.

Troubleshooting

1. The system did not recognize any face in the uploaded image.
   Please make sure you have uploaded a face image. If so, try uploading a better version of it. If the image is taken by the camera of a mobile device, make sure the resolution settings are high enough.

2. The number of hits in the desired confidence level is smaller than number of results asked by the user.
   Try lowering the certainty level. The system will suggest a certainty level that is sufficient for the requested number of identities.

3. The system could not find a significant facial landmark for explaining the prediction.
   Try changing the GRAD-CAM threshold. It is possible that for a specific query image the algorithm will not return a facial landmark on any threshold.