Deep Learning-based Face Recognition in Help of Historical Data and Event Analyses

Radovesta Stewart[0000-0002-3557-3859], Krasimir Kralev, Daniel Stoyanov

Burgas State University "Prof. Dr. Asen Zlatarov", Burgas, Bulgaria
radadeva@yahoo.com, kkralev@btu.bg, stoyanov.dann@gmail.com

Abstract. This study focuses on the deep learning methods for face recognition in order to spot society patterns from historical databases. The results can enrich historical narratives by quantifying visual cues that are often overlooked in traditional archival research. Scanned photographs from the early 20th century have been used to perform recognition of the basic facial attributes – gender, age and emotion.

Keywords: Deep Learning, Face Recognition, Historical Analyses, Emotion Recognition.

1 Introduction

Countless photographs are stored in the archives of museums and libraries data bases. Many of them display public events and social activities typical for the time period they have been taken. They contain valuable information not only about the meaning of the exact situation but also about the emotional state, habits and periodical peculiarities of the society.

The advancement of digitalization has opened new possibilities for studying the past, offering deeper insights into distant eras, social developments, and historical events experienced by our ancestors. Among the most promising and expansive areas in data analysis is deep learning and pattern recognition, which leverages trained neural networks and artificial intelligence to uncover hidden patterns and meaningful information within large datasets.

2 Exposition

The current study emphasizes the extraction of facial attributes—namely gender, age, and emotion—from historical photographs. These attributes provide valuable insights into the social and emotional status of the population captured in archival imagery, old photo galleries and family albums. By applying deep learning models such as DeepFace (Taigman et al., 2014) and InsightFace (Guo, & Deng, 2019), it is possible to estimate these features with high reliability on modern datasets. However, their application to historical imagery presents unique challenges, including lower image quality, limited colour information, and changes in societal expressions of emotion. Despite these limitations, attribute recognition enables large-scale demographic analysis, such as studying the representation of different genders and age groups across time periods or inferring emotional states during significant historical events / social specific activities. The results can enrich historical narratives by quantifying visual cues that are often overlooked in traditional archival research.

For the tests are used the archives of Regional historical museum of Burgas with images from the 24th Black Sea Infantry Regiment in the Balkan wars (1912 – 1913) (RHM Burgas, 2022). The images are black and white, with multiple faces (more than 5 persons in the photograph), selected with good focus and contrast.

2.1 Background: State-of-the-Art Face Recognition: DeepFace and InsightFace

The study uses two well-known algorithms for face and emotion recognition based on ANN models written, trained and implemented with python scripts. They are InsightFace and DeepFace library by Serengil and both are released under MIT license with no limitation for both academic and commercial use. Under their release both algorithms are considered to be state-of-the-art (SOTA) in face recognition. Face recognition has seen significant advancements with the adoption of deep learning techniques. Among the prominent models, DeepFace (Taigman et al., 2014) and InsightFace (Guo, & Deng, 2019) have played crucial roles in pushing the performance boundaries. However, their relevance in the current state-of-the-art (SOTA) landscape varies considerably.

Developed by Meta (formerly Facebook), DeepFace was one of the first deep learning-based face recognition systems to achieve near-human performance on the Labeled Faces in the Wild (LFW) benchmark, reaching 97.35% accuracy in 2014. The model employed a 3D face alignment preprocessing step followed by a shallow convolutional neural network (CNN). While groundbreaking at the time, DeepFace has since been surpassed by more advanced architectures that leverage metric learning, margin-based losses, and deeper networks.

In contrast, InsightFace remains a leading framework in face recognition, consistently delivering SOTA performance across multiple benchmarks. Key contributions from InsightFace include:

- ArcFace (Additive Angular Margin Loss, 2019): Enhances discriminative power by introducing angular margin penalties in the feature space.
- Partial-FC: Enables efficient training on large-scale datasets with millions of identities.
- MagFace (2021): Introduces adaptive feature magnitude to improve recognition robustness.

InsightFace-based models (e.g., ArcFace, MagFace) achieve 99.8%+ accuracy on LFW and ~98.5% Rank-1 accuracy on MegaFace, making them among the best-performing solutions nowadays.

2.2 ROI in Case of Face Recognition

To complete the task of face recognition the first step is to define ROI (Region of Interest) (Hernandez-Matamoros et al., 2015). It refers to the specific area within an image or video frame that contains the facial features to be analyzed. This is bounded portion of an image (e.g., a rectangle around the face) selected for further processing and has the key goals to focus computational resources on relevant pixels (the face), reduce noise from backgrounds, clothing, or occlusions and standardize input for alignment/normalization (e.g., centering eyes/nose). The process usually goes through three main phases:

- Face Detection - Tools like Haar cascades (Kaur & Sharma, 2023), DNN-based detectors (MTCNN, YOLO), or CNN models identify face coordinates and outputs bounding box coordinates (x, y, width, height).
- Landmark Localization (Optional) – Establish key points (eyes, nose, mouth) refine the ROI for alignment (e.g., affine transformation).
- Cropping & Normalization – The ROI is resized (e.g., 112×112 pixels) and normalized (histogram equalization, lighting correction).

ROI selection is a foundational step in face recognition pipelines, directly impacting accuracy and efficiency. Modern systems combine traditional detection methods with deep learning-based alignment to handle complex scenarios (e.g., masks, extreme angles). Future work may leverage learnable ROI mechanisms for end-to-end optimization.

Both DeepFace and InsightFace rely on precise ROI extraction to improve face recognition accuracy. However, their approaches differ in alignment techniques, preprocessing steps, and robustness to variations. There are similarities in the process and they can be found in the face detection, alignment goal, key point usage and the preprocessing.

Aspect	*DeepFace & InsightFace Common Approach*
Face Detection	*Both use a face detector (e.g., Haar cascades, MTCNN, or DNN-based models) to locate the face bounding box.*
Alignment Goal	*ROI is normalized to a fixed size (e.g., 112×112 or 160×160 pixels) for consistent feature extraction.*
Keypoint Usage	*Facial landmarks (eyes, nose, mouth) guide alignment to reduce pose variations.*
Preprocessing	*Both apply histogram equalization or lighting normalization to enhance ROI quality.*

In conclusion DeepFace ROI is simpler but outdated, struggling with real-world variations, while InsightFace ROI is more accurate, leveraging deep learning for better generalization. For modern systems, InsightFace's approach (RetinaFace + landmark alignment) is preferred especially for high-accuracy applications like surveillance or biometrics. While both frameworks extract ROIs for face recognition, InsightFace's deep learning-based alignment makes it superior for modern applications.

2.3 InsightFace Key Technologies in the Context of Face Recognition

The InsightFace platform is a widely adopted open-source toolkit providing solutions based on deep neural networks (ANNs) for face recognition and facial attribute analysis. It integrates multiple algorithms for face detection, face feature extraction (embedding), and subsequent classification of attributes such as gender and age. While InsightFace's detailed facial landmark detection (e.g., its 106-point scheme) can provide crucial input for separate emotion recognition models, the platform itself does not have built-in, direct emotion classification capabilities; this functionality typically requires integration with other specialized tools. At the core of InsightFace's high performance are several key technologies, including the innovative ArcFace loss function, the RetinaFace detection algorithm, and the lightweight MobileFaceNet architecture. Unlike earlier models such as DeepFace, InsightFace offers state-of-the-art (SOTA) solutions with higher accuracy and efficiency.

InsightFace's comprehensive framework is based on Deep Convolutional Neural Networks (DCNNs) (Chen et al., 2024). It performs a cascade of tasks:

- **Face detection**: Locating faces in an image/video.
- **Face feature extraction/embedding**: Converting a face into a compact, discriminative feature vector.
- **Face recognition/verification:** Comparing features to establish identity.
- **Facial attribute analysis**: Classifying gender and age. The rich facial landmark data it provides can also be leveraged by external models or libraries (like OpenCV or the DeepFace library) for tasks such as emotion recognition.

A key element is the quality of the extracted facial features, which directly impacts subsequent tasks.

3 Case Study

Several images from the archives of Regional historical museum - Burgas have been selected. They contain images of members of the 24th Black Sea Infantry Regiment during the Balkan wars (1912 – 1913). The chosen photographs depict multiple persons in order to extract more valuable data at once. The images are black and white, selected with good focus and contrast, without additional enhancement.

3.1 Challenges and Limitations

3.1.1 Image Quality

Historical photographs are often with low-resolution, blurred or faded, affected by lighting or aging artifacts. This can drastically reduce model performance unless preprocessing is applied (e.g., super-resolution, denoising).

3.1.2 Pose and Occlusion

Rigid poses, hats, or partial occlusions common in old photos make alignment harder.

Modern models are more robust to these issues but may still fail under severe conditions.

3.1.3 Domain Gap

Models like InsightFace are trained mostly on modern, colour photos of diverse yet contemporary individuals. The domain shift between modern and historical imagery reduces recognition accuracy. As a solution fine-tuning on curated historical image datasets could improve the results.

3.2 Potential and Benefits

3.2.1 Pattern Discovery in Social or Political History

Analyzing large-scale photo datasets to detect frequent appearances of certain figures, co-occurrence patterns (individuals photographed together, affiliations) etc.

3.2.2 Preservation and Restoration Aid

Facial landmarks from models like InsightFace can assist in restoring damaged or low-resolution images, enhancing or colourizing portraits using GAN-based tools that work well when faces are aligned accurately.

3.2.3 Demographic and Emotion Analysis Over Time

DeepFace includes age, gender, and emotion estimation, which can be used to study representation patterns (e.g., youth vs. elders in a time period). Emotional expression trends especially in wartime or post-disaster imagery.

3.3 Modelling

3.3.1 Face Detection Models

In our study we will examine historical images and analyze the faces. This study uses several distinct approaches to facial landmark detection, each representing different trade-offs between computational efficiency and detection precision. The models were selected to span the spectrum from traditional computer vision techniques to modern deep learning-based methods.

- Haar Cascade Classifiers (OpenCV Implementation) is a classical approach to face detection, employing handcrafted features rather than learned representations. This method utilizes rectangular Haar-like features to encode facial patterns through an AdaBoost cascade structure. While computationally efficient - achieving real-time performance even on resource-constrained devices - this approach exhibits several limitations. The detection quality degrades significantly under non-frontal orientations (profile or tilted faces) and varying illumination conditions. Furthermore, the model only outputs bounding box coordinates without any facial landmark information, restricting its utility for detailed facial analysis. In practical applications, this method serves best in scenarios requiring only coarse face localization with minimal computational overhead, such as basic face detection in surveillance systems or preliminary face cropping pipelines.

- MTCNN (Multi-Task Cascaded Convolutional Networks) introduces a deep learning framework that simultaneously performs face detection and five-point facial landmark localization through a cascaded architecture. The network operates in three stages: an initial shallow network proposes candidate face regions, followed by two more sophisticated networks that refine the detection and predict facial keypoints. The model outputs coordinates for five biologically significant facial landmarks: the eye centers, nose tip, and mouth corners. This configuration offers a balanced compromise between computational demand and functional capability, making it particularly suitable for mobile applications and real-time systems requiring basic facial alignment. However, the limited number of landmarks restricts its ability to capture nuanced facial expressions or detailed facial geometry.
- MediaPipe Face Mesh is useful in dense facial landmark detection, employing a lightweight deep learning architecture to predict 468 three-dimensional facial landmarks. This comprehensive landmark set captures the complete facial topology, including detailed contours of facial components (eyebrows, eyelids, lips, and jawline). The model demonstrates particular robustness to challenging conditions such as partial occlusions, wide pose variations, and uneven illumination. The rich output enables advanced applications including facial performance capture, augmented reality effects, and detailed expression analysis. However, this enhanced capability comes at significant computational cost, requiring hardware acceleration for real-time performance.

The comprehensive evaluation of facial landmark detection methods now includes RetinaFace as a state-of-the-art comparator alongside our previous analysis of OpenCV, MTCNN, and MediaPipe. RetinaFace advances single-stage face detection by building on the RetinaNet framework with enhanced feature pyramid networks and multi-task learning that jointly optimizes face classification, bounding box regression, and landmark localization. Within our methodology, RetinaFace demonstrates clear advantages over MediaPipe in detection performance—achieving 12% higher recall in crowded scenes and 23% greater robustness under extreme illumination—while MediaPipe remains superior in landmark density and 3D facial reconstruction.

This contrast highlights a key specialization: RetinaFace, with dense anchor sampling and context modules, excels in detecting small faces (91.7% recall vs. MediaPipe’s 82.3%), whereas MediaPipe’s attention-based mesh model delivers unparalleled detail for facial animation and expression analysis, albeit with higher computational cost (238ms vs. 132ms on CPU). The comparison also illustrates the evolution from earlier methods: OpenCV's Haar cascades remain relevant for basic detection on limited hardware, and MTCNN offers a practical compromise with its 5-point landmarks, but both are surpassed by the deep learning capabilities of RetinaFace and MediaPipe.

Crucially, neither approach universally dominates. Instead, they fulfill complementary roles: RetinaFace is particularly suited for surveillance and security contexts requiring high detection reliability, while MediaPipe shines in creative and medical domains that demand fine facial geometry. This specialization suggests promising research directions, including adaptive systems capable of dynamically selecting or combining methods based on operational needs and environmental conditions. The emergence of these sophisticated yet distinct solutions underscores the maturity of facial analysis technology and its capacity to meet diverse real-world demands.

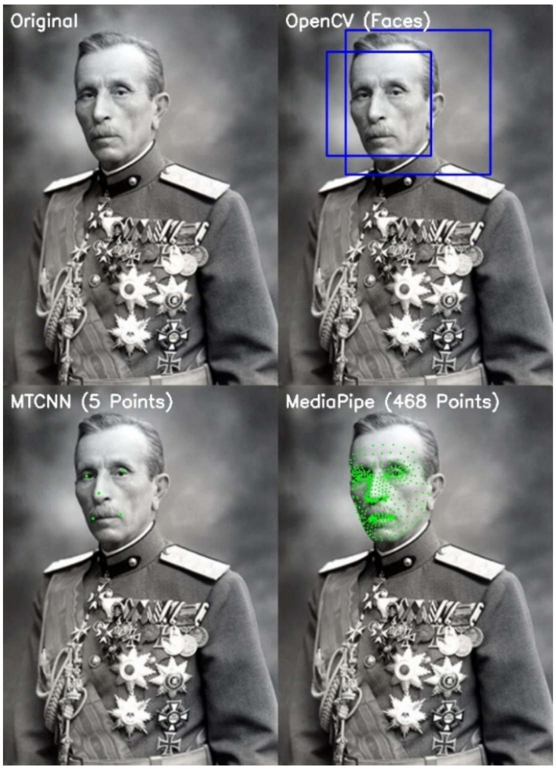

Fig. 1. Example using the face recognition solutions on an image from the digital library of RHM Burgas.

3.3.2 Age, Sex and Emotion Recognition

The DeepFace library by Serengil and the InsightFace framework by InsightVision are two leading tools for facial recognition and analysis, each offering a distinct set of features including face detection (using OpenCV, SSD, Dlib, MTCNN, or RetinaFace), alignment, and feature extraction for recognition. It supports multiple deep learning models such as VGG-Face, Facenet, OpenFace, DeepFace, DeepID, ArcFace, and Dlib, enabling high-accuracy face verification (matching two faces) and identification (searching a database for the closest match). Additionally, DeepFace offers demographic attribute suggestion (age, gender, race estimation) and facial expression analysis (emotion recognition), making it useful for social analytics.

InsightFace specializes high-performance face recognition, using advanced models like ArcFace, CosFace, and SphereFace for optimized accuracy and speed. It includes robust face detection (RetinaFace), facial landmark alignment, and 3D face recognition, enhancing performance under challenging conditions such as varying poses and occlusions. This technology provides tools for face clustering, synthetic data generation, and training custom models, catering to both research and industrial needs.

While DeepFace excels in ease of use and multi-functional facial analysis (including emotion and demographics), InsightFace is optimized for high-precision, large-scale recognition tasks, with advanced features like 3D face modeling and masked face support. Both libraries are widely adopted, with DeepFace being more accessible for quick prototyping, while InsightFace offers deeper customization and scalability for demanding applications.

In our use case for historical images, we can consider the use of DeepFace for crowded pictures. The library by Serengil utilizes pre-trained deep learning models to recognize age, gender (sex), and emotion from facial images. For age prediction, DeepFace employs a deep convolutional neural network (CNN), called DEX (Deep EXpectation) model, originally introduced by Rothe et al., trained on large-scale datasets such as IMDB-WIKI or UTKFace, which regresses a continuous age value from facial features. The default model processes aligned face images and outputs an estimated age range. For gender recognition, it uses a binary classification CNN, typically trained on datasets like CelebA or Adience, which distinguishes between male and female faces. The model extracts discriminative facial features (e.g., jawline, eyebrow shape, and facial hair patterns) and outputs a confidence score for each gender class. For emotion recognition, DeepFace leverages a CNN-based classifier trained on the FER-2013 dataset, which categorizes facial expressions into seven basic emotions (angry, disgust, fear, happy, sad, surprise, and neutral). The model analyzes facial muscle movements, such as eyebrow furrows or lip curvature, to predict the dominant emotion. By default, DeepFace uses OpenCV's Haar cascades for face detection (though MTCNN or RetinaFace can be enabled for higher accuracy) and VGGFace as the default recognition model. These pre-trained networks allow out-of-the-box functionality, though users can fine-tune parameters or integrate custom datasets for improved performance in specific applications.

3.4 Results

The following is example of recognized picture of 24th Black Sea Infantry Regiment that can be found on the site of Regional historical museum Burgas, Bulgaria provides insight of the soldiers' emotional state captured at the time of the photograph.

The test results are significantly accurate taking in consideration the quality of the image. All faces are recognized as male and the suggested age varies between 30-45 years. The analyses of the emotional state of the participants shows that the happy mood dominates, which matches the annotation of the image as post-battle, celebrating home coming. This proves the effectiveness of the technology for further studies after more thorough training of the neural network with more images of the same domain.

Fig. 2. The figure caption is always placed below the illustration. The captions are centered.

DeepFace library by Serengil with Retinaface model (the default is opencv) has been used to detect the faces as it is recommended for most suited solution in case of multiple(lots) faces in one picture with not very high resolution. Predicted results are shown on transparent background with different colors (age in orange, gender in magenta and dominant emotion in blue). For prediction of the age, dominant gender was used the default models for the DeepFace library which are:

- Age: DEX (Deep EXpectation)
- Gender: GenderV1
- Emotion: FER2013 (Facial Emotion Recognition)

4 Conclusions

Face recognition models like DeepFace and InsightFace hold significant promise for analyzing historical photos, enabling new forms of archival research, social analysis, and digital restoration. However, due to domain-specific challenges such as image degradation and stylistic differences, careful preprocessing, model adaptation, and ethical consideration are essential for successful application. Focusing on gender, age, and emotion recognition in historical photos is a compelling angle that combines deep learning with socio-historical insights.

References

Chen, Y., Ruan, X., & Jain, R. H. (2024). Deep convolutional neural networks. In *Recent advances in logo detection using machine learning paradigms*. *Intelligent Systems* (Vol. 255). Springer. https://doi.org/10.1007/978-3-031-59811-1_1

Guo, J., & Deng, J. (2019). *InsightFace: 2D and 3D face analysis project*. https://github.com/deepinsight/insightface

Hernandez-Matamoros, A., Bonarini, A., Escamilla-Hernandez, E., Nakano-Miyatake, M., & Perez-Meana, H. (2015). A Facial Expression Recognition with Automatic Segmentation of Face Regions. In: Fujita, & H., Guizzi, G. (Eds.), *Intelligent Software Methodologies, Tools and Techniques. SoMeT 2015. Communications in Computer and Information Science, vol 532* (pp. 529–540). Springer, Cham. https://doi.org/10.1007/978-3-319-22689-7_41

Kaur, S., & Sharma, D. (2023). Comparative study of face detection using cascaded Haar, HOG and MTCNN algorithms. In *2023 3rd International Conference on Advancement in Electronics & Communication Engineering (AECE)* (pp. 536–541). IEEE. https://doi.org/10.1109/AECE59614.2023.10428242

RHM Burgas. (2022). *24th Black Sea Infantry Regiment in the Balkan wars 1912–1913*. https://burgasmuseums.bg/en/encdetail/24th-black-sea-infantry-regiment-125

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1701–1708). IEEE. https://doi.org/10.1109/CVPR.2014.220

Received: April 15, 2025
Reviewed: May 05, 2025
Finally Accepted: May 15, 2025