

Designing a Face Recognition System for Event Photography

Nicholas Borges de Vasconcelos, Mel Bozzola, Adam Bennet Bleiel, Rigel P. Fernandes[✉], Victor A. Azevedo Costa Santos, and Clayton J. A. Silva

Abstract—Event photography produces large collections of images that are difficult to search manually, limiting participants’ ability to retrieve photographs in which they appear. This work presents a face-based image retrieval system designed for this context, combining a ResNet face encoder with Euclidean distance search over a pre-indexed gallery of embeddings. The system was evaluated on SBrT2025, a real-world dataset of 1,427 images collected at a Brazilian telecommunications conference, where 81.2% of all detected faces were classified as small faces, a challenge characteristic of event photography. Manual annotation of 714 images (50% of the dataset) shows that the pipeline achieves a face-presence recall of 88.5%. Detection accuracy varied significantly by image composition, reaching 66.8% exact match on solo images but dropping to 35.3% on group shots, where occlusion and small face scale remain the primary sources of error. These results establish a detection-accuracy baseline for face-based retrieval in event photography and motivate future work on identity-aware evaluation and retrieval robustness under unconstrained real-world conditions.

Keywords—Face Recognition, Image Retrieval, Event Photography, Person Search, Small Faces.

I. INTRODUCTION

Event photography often produces large collections of images containing many different attendees across several days. Thousands of photographs are taken simultaneously, generating an enormous volume of images that need to be organized, cataloged and distributed to the correct participants in an efficient manner. Manually searching these collections to find images of a specific person can be slow and inefficient, especially when the photographs include a mixture of portraits, group photos, audience scenes, and informal interactions [1].

It is in this scenario that facial detection technology presents itself as a strategic tool capable of transforming the way event photographs are organized and delivered, enabling each participant to automatically receive only the images in which they appear, without the need for manual intervention [2]. Computer vision emerges as a promising response to this problem, especially through facial recognition systems based on machine learning. This technology allows computers to interpret and analyze images in real time, identifying human faces automatically, quickly and without the need for

contact or active cooperation from the user [2]. In recent years, the advancement of deep learning algorithms, especially convolutional neural networks, has significantly increased the accuracy of these systems, making them comparable to or even superior to human recognition capability under certain conditions [3]. At the same time, the popularization of open source libraries such as OpenCV and Dlib has reduced the technical and financial barriers to the development of these solutions, making their application viable in smaller corporate contexts [4]. Given this scenario, the present work implements and evaluates a person retrieval system using computer vision for facial recognition applied to events, analyzing its technical and operational viability.

II. BACKGROUND

Face recognition and person retrieval systems rely on the extraction of visual features that can distinguish one individual from another. Earlier approaches used handcrafted facial descriptors and geometric relationships between facial components, but modern systems usually rely on deep convolutional neural networks to learn discriminative feature representations directly from image data. The development of residual networks showed that very deep architectures could be trained effectively through identity shortcut connections, making deep feature extraction more reliable for image-recognition tasks [5]. This idea is important for modern recognition systems because faces, pedestrians, and other visual targets are commonly represented as compact embedding vectors that can be compared in a learned feature space.

This work is closer to face-based image retrieval than to conventional identity verification: rather than deciding whether two images depict the same person, the system ranks a gallery of candidates by similarity to a query face. This formulation follows the person search paradigm [6].

Although many person-search and re-identification methods focus on full-body pedestrian retrieval, the present work studies a related problem in event photography using faces as the primary visual cue. This setting introduces practical challenges that are also discussed in the person-search literature, especially the dependence of retrieval quality on detection and localization. Poor bounding boxes can reduce the quality of the extracted representation, and previous work has shown that refining localization for re-identification can improve retrieval performance [7]. In the SBrT2025 dataset, this issue is particularly relevant because many images contain group scenes, small faces, partial occlusions, and variable

Nicholas Borges de Vasconcelos, Mel Bozzola, Rigel P. Fernandes, Victor A. Azevedo Costa Santos, and Clayton J. A. Silva are affiliated with the Computer and the Software Engineering programs at the (Ibmec University Center, Rio de Janeiro, Brazil). Emails: nicholas@gmail.com, bozzolamel@gmail.com, rigelfernandes@gmail.com, clayton.silva@ibmec.edu.br.

Adam Bennet Bleiel is affiliated with the General Management program at the Cologne Business School (CBS), Cologne, Germany. Email: bleiel@gmail.com.

camera-to-subject distances and multi-scale variation across audience, solo, and group shots.

III. THE METHOD

A. System Overview

The proposed system, shown in Fig. 1, begins with a query image (either a selfie or a registration photo) and proceeds to detect the largest face in the image. That image is then cropped and processed using 128-dimensional embedding. The resulting vector representation is compared against the embeddings stored in the indexed gallery. The result is a ranked list of the matching images based on the computed distances of the face query embedding against all pre-indexed gallery face entries.

Because detection and embedding are computationally expensive, the full gallery was processed offline before retrieval. Query-time search then requires only embedding the query face and comparing it against the pre-indexed gallery of 3,344 face entries via linear scan.

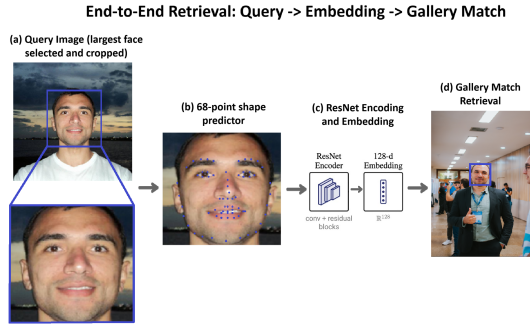


Fig. 1: End-to-end retrieval pipeline. The query face is detected and embedded; the gallery is then searched for all indexed embeddings within Euclidean distance $\tau = 0.60$, and results are ranked by ascending distance.

B. Face-Embedding

For face detection and embedding, the Python library *face_recognition*[8] is used. It wraps the dlib ResNet face recognition model for ease of use. Given an input image, the face detector identifies all faces in the image and crops the largest face using the retrieved bounding box. The cropped image is then processed using a 68-point shape predictor to estimate facial landmarks, followed by a ResNet encoder that produces a 128-dimensional floating-point descriptor.

Similarity between the query and the gallery embeddings is measured by Euclidean distance, defined as:

$$d_f(q, g) = \|e_q - e_g\|_2 \quad (1)$$

where e_q is the query face embedding and e_g is the gallery face embedding. A gallery entry is considered a match if its face distance falls below a certain threshold τ_f .

$$d_f(q, g) < \tau_f \quad (2)$$

For this project the threshold has been set to 0.60 by default, the recommended operating point for the dlib model.

C. Dataset

The dataset used in this work was built from official photography of the SBtT2025 event, held in Salvador, Brazil, covering four conference days and containing 1,427 images. The images were captured by a professional photographer under real conference conditions, this includes presentations, audience scenes, informal interactions, and group photographs, and were not collected for biometric evaluation purposes. This distinguishes the collection from controlled face-recognition datasets and introduces challenges typical of event photography: pose variation, mixed lighting, motion blur, and wide camera-to-subject distances.

The collection contains 1,427 images, of which 1,219 yielded a total of 3,344 detected faces (an average of 2.34 faces per image). The bounding-box area highlights a skewed distribution of face scales: while the mean area is 167,531 px, the median is only 72,361 px, emphasizing the prevalence of small, distant faces.

The dataset presents challenges that motivate its use as an event-photography retrieval benchmark. Face scale varies substantially: the mean bounding-box area was nearly twice the median, indicating a skewed distribution driven by a small number of close-up portraits while the majority of detections correspond to small, distant faces. At a threshold of 1% of image area, 81.2% of all detections were classified as small faces. These characteristics differ markedly from controlled person-search benchmarks such as CUHK-SYSU [6]. The SBtT2025 dataset instead reflects the unconstrained nature of event photography, where group composition, small faces, and occlusion are central sources of difficulty.

IV. RESULTS

The evaluation reported to follow characterizes the detection and counting performance of the pipeline. To evaluate face-count accuracy, 714 images were manually detected and annotated with ground-truth face counts, covering approximately 50% of the full dataset sampled randomly across four days. A face was considered detectable if more than half of their face was visible and not influenced with a heavy amount of blur. The signed error (detected – ground truth) was recorded for each image.

A. Detection Accuracy by Category

Figure 2 shows the exact match rate and mean absolute error per composition class. The detector performed substantially better on solo images (66.8% exact match, MAE = 0.40) than on group images (35.3% exact match, MAE = 1.24). The *no face detected* class achieved the highest exact match rate (78.6%, MAE = 0.31), confirming most as true negatives. Overall, the pipeline matched ground-truth counts in 367 of 714 images (50%), with MAE = 0.83. Table I reports the full breakdown.

B. Error Distribution

Figure 3 shows the distribution of signed detection errors across the annotated subset. The distribution is markedly

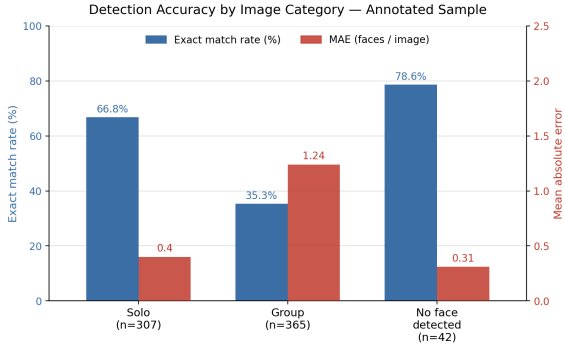


Fig. 2: Detection accuracy per image composition class (714-image annotated subset).

TABLE I: Face-count detection accuracy on the 714-image annotated subset. Exact = detected count equals ground truth; Over = detected > ground truth; Under = detected < ground truth.

Class	N	Exact	Exact %	Over	Under	MAE
Solo	307	205	66.8%	39	63	0.40
Group	365	129	35.3%	34	202	1.24
No Face	42	33	78.6%	9	0	0.31
Total	714	367	51.4%	82	265	0.82

asymmetric: under-detection accounted for 265 of the 347 total errors (76.1%), while over-detection accounted for the remaining 82 (23.6%). The most frequent non-zero error was -1 , occurring in 87 images, indicating that the most common failure mode is missing exactly one face. Errors beyond ± 3 were rare, appearing in fewer than 3% of images.



Fig. 3: Signed detection error distribution (detected – ground truth) across the annotated subset; under-detection is the dominant failure mode.

C. Limitations of Count-Based Evaluations

While exact face count is a natural metric, it fails to capture coincidental matches. As shown in Figure 4, the detector identified a printed banner face (false positive) but missed the foreground attendee (false negative). This yields a net-zero error for the image, masking a complete retrieval failure. Although these cases are not frequent they highlight a fundamental limitation of the count-based evaluation and motivates the use of identity-aware ground truth.



Fig. 4: Example of a false positive and a false negative.

V. CONCLUSION

This work presented a face-based image retrieval system designed for event photography, evaluated on a real-world dataset of 1,427 images collected at SBt2025. The system combines a dlib ResNet face encoder with Euclidean distance search over a pre-indexed gallery, and was deployed as a web application accessible to conference participants. Manual annotation of 714 images (50% of the dataset) revealed that the pipeline achieves a 51.3% exact face-count match rate, with substantially higher accuracy on solo images (66.8%) than on group shots (35.3%). Under-detection was the dominant failure mode, driven primarily by the high proportion of small faces inherent to event photography. The end-to-end retrieval pipeline demonstrated qualitatively correct results for frontal, well-lit query faces, while remaining sensitive to distance threshold selection and vulnerable to false positives from visually similar embeddings. Future work includes the adoption of identity-aware ground truth to enable precision and recall evaluation at the retrieval level, and the integration of complementary visual cues to improve robustness.

REFERENCES

- [1] D. Motwani, A. Tyagi, V. Dabhi, and H. K. Prajapati, “Automated attendee recognition system for large-scale social events or conference gathering,” in *2025 10th International Conference on Signal Processing and Communication (ICSC)*, 2025, pp. 283–288.
- [2] A. Kapiton, N. Kononets, V. Mokliak, V. Onipko, S. Dudko, V. Pylypenko, and A. Sokil, “Tecnologia biométrica de reconhecimento pessoal,” *Revista on line de Política e Gestão Educacional*, vol. 28, no. 00, p. e023015, 2024.
- [3] L. Costa, T. Sousa, F. Silva, L. Almeida, D. Pereira, A. Artero, and M. Piteri, “Análise de métodos de detecção e reconhecimento de faces utilizando visão computacional e algoritmos de aprendizado de máquina,” *COLLOQUIUM EXACTARUM*, vol. 13, pp. 01–11, 09 2021.
- [4] W. d. Oliveira, “Identificar autoridades por meio de reconhecimento facial: uso de tecnologia de visão computacional como alternativa para antigo processo de fotogramas (Carômetro),” Trabalho de Conclusão de Curso, Centro Universitário de Brasília (Uniceub), 2020, orientador: William Roberto Malvezzi. [Online]. Available: <https://repositorio.uniceub.br/jspui/handle/prefix/15431>
- [5] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [6] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “Joint detection and identification feature learning for person search,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 3415–3424.
- [7] C. Han, J. Ye, Y. Zhong, X. Tan, C. Zhang, C. Gao, and N. Sang, “Re-id driven localization refinement for person search,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 9814–9823.
- [8] A. Geitgey, “Face recognition: The world’s simplest facial recognition api for python and the command line,” https://github.com/ageitgey/face_recognition, 2018.