Denoising Face Recognition via DBSCAN

Trung Vu, Sihyun Lee, Alexandre Lamy

Department of Computer Science, Columbia University

COMS 4995: Methods in Unsupervised Learning (Fall 2018)

COLUMBIA COLLEGE
COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

COLUMBIA ENGINEERING
The Fu Foundation School of Engineering and Applied Science

ABSTRACT

In this paper we address the problem of detecting and recognizing specific objects (such as the faces of a specific group of people) in real life videos. By using unsupervised clustering techniques, we are able to exploit the inter-frame relationships in the video to significantly improve the accuracy of basic object detection and recognition algorithms. This yields results which are also much more robust to frame switches, lengthy occlusions, and variable numbers of targeted objects leaving and reentering the frames (all common occurrences in real life videos) than standard tracking techniques.

MOTIVATION

Potential solutions:

Tracking:
- Common algorithms: OLB, IVT, MIL, L1, TLD
- They all make strong assumptions on video structure (inter frame relationships)
- Cannot handle frame switches, objects coming in and out, and long occlusions

Frame by frame object detection/recognition:
- No assumptions concerning the relations between different frames: more robust yet does not capture interframe dependencies
- Very noisy and vulnerable to subtle differences

Our proposed solution:

Use frame by frame object recognition/detection methods but denoise those results using a simple post processing method which exploits much weaker assumptions on video structure.
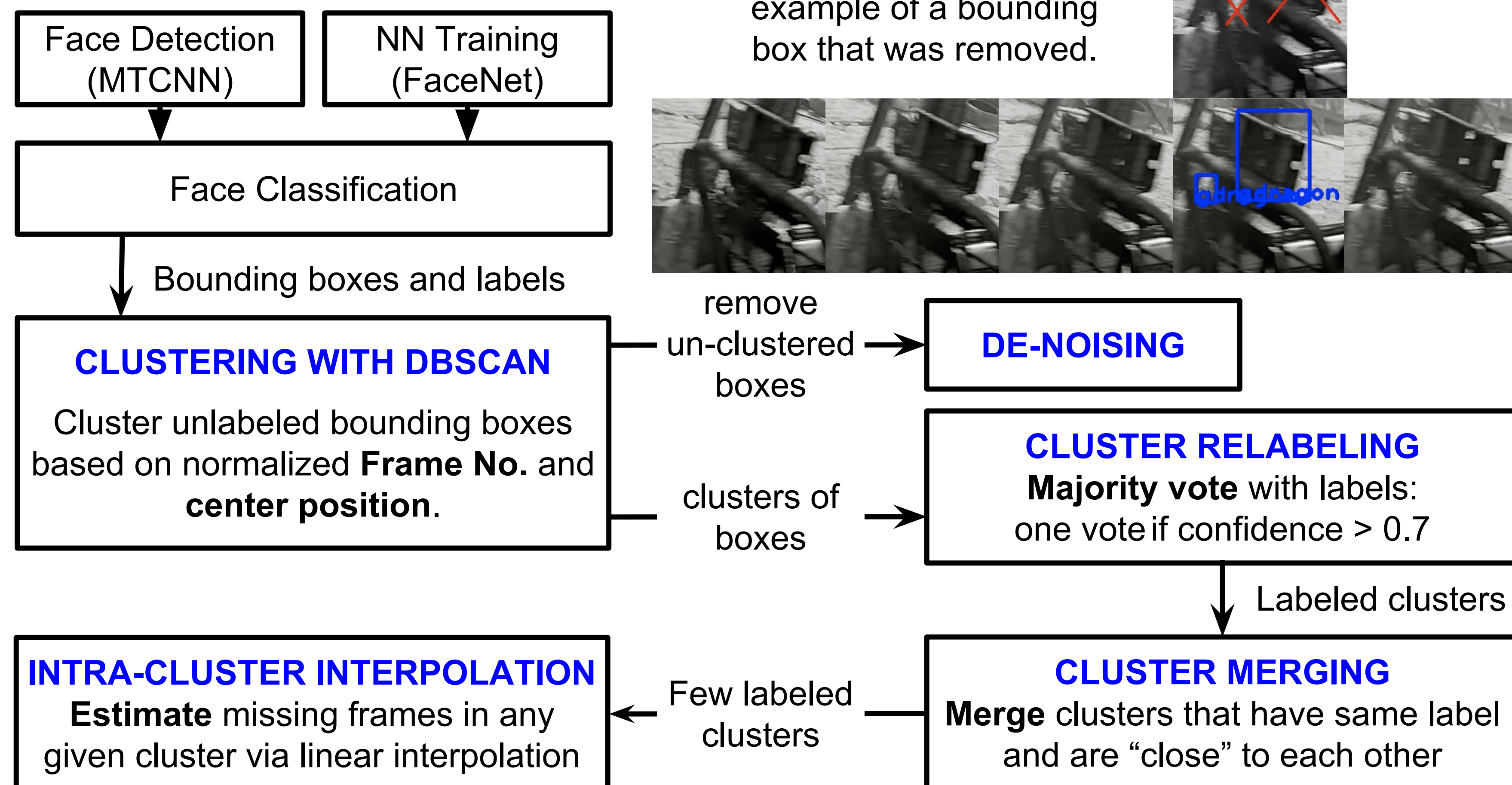Yields better and more robust performance.

METHODS & RESULTS

Face Detection (MTCNN) → Face Classification

NN Training (FaceNet) → Face Classification

Face Classification → Bounding boxes and labels

CLUSTERING WITH DBSCAN
Cluster unlabeled bounding boxes based on normalized **Frame No.** and **center position**.

remove un-clustered boxes → **DE-NOISING**

clusters of boxes → **CLUSTER RELABELING**
Majority vote with labels: one vote if confidence > 0.7

→ Labeled clusters →

CLUSTER MERGING
Merge clusters that have same label and are "close" to each other

Few labeled clusters → **INTRA-CLUSTER INTERPOLATION**
Estimate missing frames in any given cluster via linear interpolation

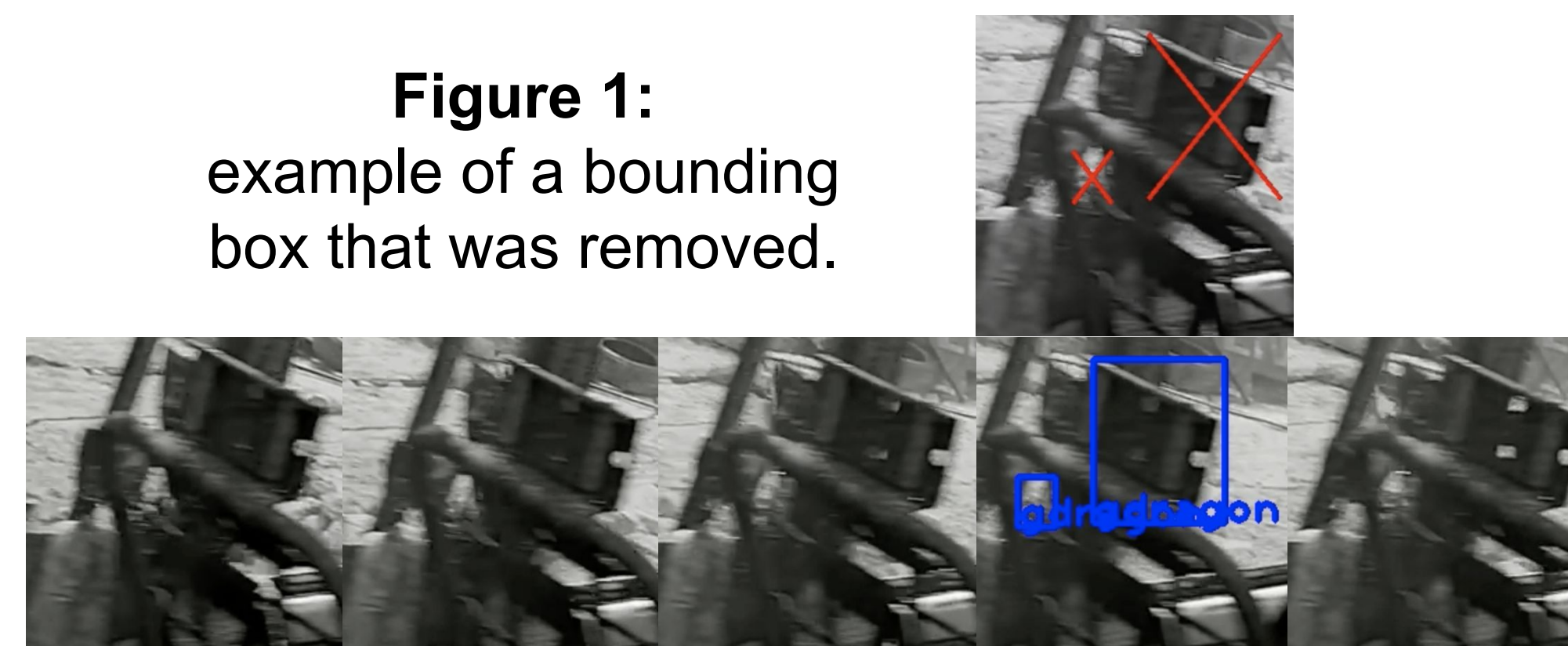**Figure 1:** example of a bounding box that was removed.

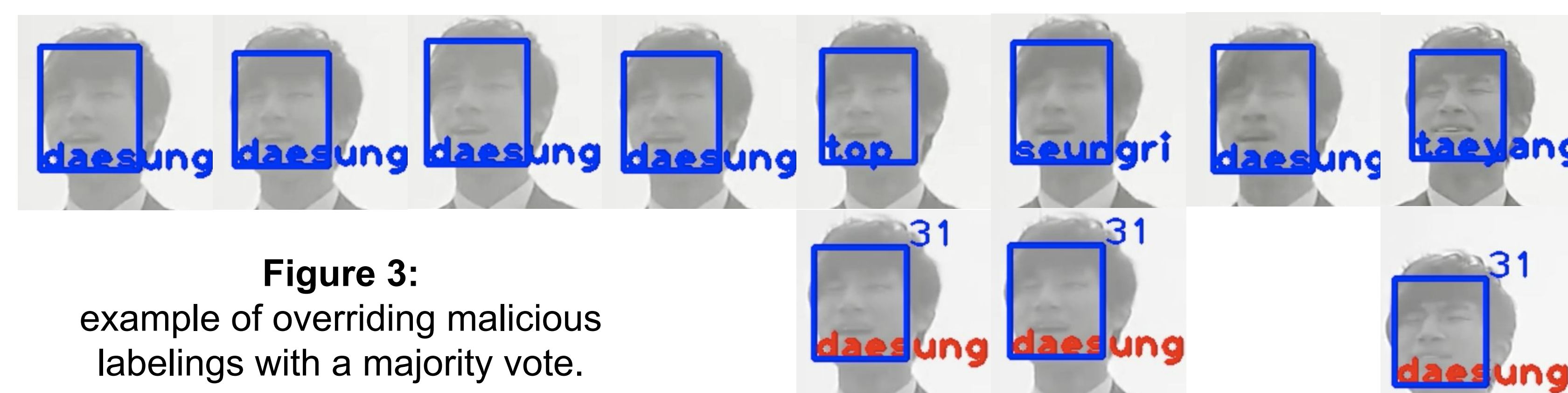**Figure 2:** example of tracking a face in presence of occlusion.

Figure 3: example of overriding malicious labelings with a majority vote.

CONCLUSION

Pros of our Algorithm:
- Simple and efficient
- Applies to all detection/recognition algorithms
- Fewer assumptions than tracking methods so more robust (can handle occlusions, people moving in and out, and scene changes)
- Correct labeling for 93% of the clusters
- 6% of the bounding boxes are new/.
- 29% of the final bounding boxes have labels corrected or interpolated from the original.

Potential Improvements:
- More rigorous techniques for merging of clusters (secondary clustering, second DBSCAN run with different parameters, etc.)
- More complex movement models for intra cluster interpolation (could consider velocity)
- Make an online version of our algorithm: speed is not a problemy but clustering and interpolation would need to happen online
- Robustness to edge cases (e.g. person moving to previous position of another before movement ends)

ACKNOWLEDGEMENTS

FURTHER RESOURCES

The model was experimented on "LOVE SONG" Music Video by BIGBANG. The detection results are in youtu.be/pzmx2Ztb6E0. The code can be found in github.com/ttv2107/UML-project/