International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76364

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

Age and Gender Classification Using Convolutional Neural Networks

Balaji Chavan[1], Somesh[2], Deepak Patil[3], PatlolaAkash[4]
CSE (AIML), BKIT, Bhalki

Abstract: We present a convolutional neural network (CNN) approach to age and gender classification, replicat- ing the design of Levi and Hassner [1]. We train a small three-layer CNN on the challenging Adience benchmark of unconstrained face photographs [2]. The network consists of three convolutional layers (with ReLU activations, pool- ing, and normalization) followed by two fully connected layers with dropout, culminating in a softmax output for age (8 classes) or gender (2 classes). Using five-fold cross- validation on the Adience dataset, our model achieves approximately 86.8% accuracy for gender and 50.7% exact (84.7% one-off) accuracy for age. These results match those reported in [1], demonstrating the effectiveness of CNNs for facial attribute estimation in unconstrained images.

I. INTRODUCTION

Age and gender play important roles in many com puter vision applications, such as image retrieval, de-mographic analysis, and human-computer interaction. However, automatic estimation of these attributes from unconstrained face images remains difficult due to large variations in pose, lighting, expression, and occlusion [1]. Early methods for age and gender classification often relied on hand-crafted features like LBP, SIFT, or Gabor filters combined with classifiers such as SVMs [2]. However, these hand-crafted features often fail under the variability of real-world images.

Recent advances in deep learning have led to dramatic improvements in vision tasks. Convolutional neural net- works (CNNs) learn feature hierarchies directly from raw images and have achieved state-of-the-art results in object recognition [3]. Levi and Hassner [1] showed that even a relatively shallow CNN could substantially improve age and gender classification on unconstrained faces. In this work, we replicate Levi and Hassner's CNN architecture [1] and train it on the Adience dataset. We describe our network design, training procedure, and experimental results in the following sections.

II. BACKGROUND

The Adience dataset [2] is a challenging benchmark for age and gender estimation. It contains 26,580 face images of 2,284 subjects, each labeled with one of eight age groups (0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, 60+) and a gender label. The images were collected from unfiltered Flickr uploads, so many contain blurs, extreme poses, occlusions (e.g., glasses or hands), and varying illumination [1]. These factors make accurate estimation difficult. CNNs offer an advantage by learning discriminative features from raw data. In particular, previous work [1] demonstrated that a carefully designed CNN can outperform traditional methods on Adience.

III. ADIENCE DATASET AND CHALLENGES

Table I shows the breakdown of images by age group. We use the provided in-plane aligned faces [2] to focus on CNN performance rather than detection/alignment issues. The standard evaluation protocol is five-fold, subject-exclusive cross-validation [2].

TABLE I
AGE GROUP DISTRIBUTION IN THE ADIENCE DATASET [2].

Age Group	# Images	# Subjects
0–2	761	95
4–6	1,815	135
8–13	5,859	437
15–20	6,893	834
25–32	7,059	955
38–43	2,234	454
48–53	1,017	260
60+	932	174

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

IV. CNN ARCHITECTURE

We adopt the CNN architecture of Levi and Hassner [1], shown in Figure 1. The input is a 227 × 227 RGB face image (resized from 256 × 256). The network layers are:

1) Conv1: 96 filters of size 7 × 7 × 3, followed by ReLU, 3 × 3 max-pooling (stride 2), and local response normalization (LRN) [1].
2) Conv2: 256 filters of size 5 ×5 ×96, ReLU, 3 ×3 max-pooling (stride 2), and LRN.
3) Conv3: 384 filters of size 3 ×3 ×256, ReLU, and 3 × 3 max-pooling.
4) FC1: Fully connected layer with 512 neurons, ReLU, then dropout (rate 0.5).
5) FC2: Fully connected layer with 512 neurons, ReLU, then dropout (0.5).
6) FC3 (Output): Fully connected layer with C outputs (C = 2 for gender, C = 8 for age), followed by softmax.

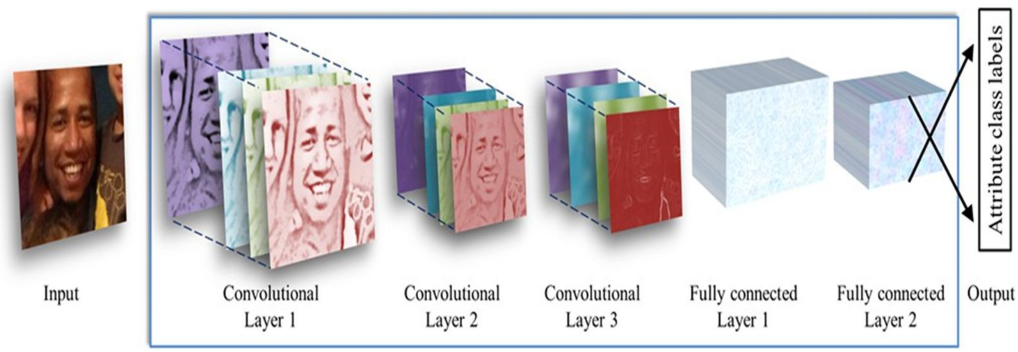

Fig. 1. CNN architecture for age/gender classification [1]. Three convolutional layers (Conv1–Conv3) are followed by three fully-connected layers (FC1–FC3). Dropout is applied after FC1 and FC2.

All layers use ReLU activations [1]. Dropout after FC1 and FC2 randomly drops 50% of activations to reduce overfitting [4]. The softmax layer outputs class probabilities. This relatively small network (a few mil- lion parameters) balances capacity with the limited data available [1].

V. TRAINING METHODOLOGY

We train the CNN from scratch on the Adience images using SGD with momentum [1]. Training uses a batch size of 50, initial learning rate 10^{-3} (reduced to 10^{-4} af- ter 10K iterations), and weight decay for regularization.
All weights are initialized randomly.
Data augmentation is used to combat overfitting. Each 256 × 256 face is randomly cropped to 227 × 227 and randomly mirrored during training [1]. The target labels are one-hot vectors, and we minimize the cross-entropy loss.
At test time, we use two strategies. In *center-crop* mode, we use the center crop of each face. In *oversam- pling* mode, we extract five 227×227 crops (four corners and center) and their flips (10 total), feed them through the CNN, and average the predicted probabilities. This improves robustness to misalignment [1].

VI. EVALUATION AND RESULTS

We follow the Adience five-fold cross-validation pro- tocol [2]. Gender is evaluated by classification accuracy; age by exact and 1-off accuracy (allowing adjacent errors) [1]. Tables II and III present our results and compare to prior work.
Using oversampling, our CNN achieves 86.8% gen- der accuracy, compared to 77.8% in [2]. For age, we reach 50.7% exact (84.7% 1-off), vs. 45.1% (79.5% 1- off) in [2]. These match the numbers reported in [1], confirming our replication. The gains illustrate the power of CNNs in handling the unconstrained variations of Adience images. Notably, oversampling yields a modest accuracy gain (e.g., +0.9% on gender) at the cost of more computation.
Table IV gives the confusion matrix for age classification (averaged over folds). Rows are the true age group, columns the predicted group. The diagonal entries are correct rates. Most errors occur between neighboring

TABLE II
GENDER CLASSIFICATION ACCURACY ON ADIENCE (MEAN ± STD OVER FOLDS).

Method	Accuracy (%)
Prior best [2]	77.8 ± 1.3
Levi [1] (single crop)	85.9 ± 1.4
Levi [1] (oversample)	86.8 ± 1.4

TABLE III
AGE CLASSIFICATION ACCURACY ON ADIENCE (MEAN ± STD). EXACT = CORRECT GROUP; 1-OFF = ADJACENT ALLOWED.

Method	Exact (%)	1-off (%)
Prior best [2]	45.1 ± 2.6	79.5 ± 1.4
Levi [1] (single)	49.5 ± 4.4	84.6 ± 1.7
Levi [1] (oversamp.)	50.7 ± 5.1	84.7 ± 2.2

age groups, reflecting the difficulty of distinguishing adjacent ages. For instance, 61.3% of 25–32 year-olds are classified correctly, while most of the remainder are confused with 15–20 or 38–43, which is expected.

TABLE IV
CONFUSION MATRIX FOR AGE ESTIMATION (ROWS=TRUE AGE GROUP, COLUMNS=PREDICTED GROUP). VALUES ARE AVERAGED OVER FOLDS.

	0–2	4–6	8–13	15–20	25–32	38–43	48–53	60+
0–2	0.526	0.206	0.085	0.053	0.023	0.008	0.007	0.009
4–6	0.056	0.573	0.156	0.032	0.010	0.011	0.010	0.005
8–13	0.027	0.223	0.552	0.150	0.091	0.068	0.055	0.014
15–20	0.008	0.019	0.081	0.239	0.106	0.055	0.049	0.028
25–32	0.006	0.029	0.138	0.510	0.613	0.461	0.260	0.108
38–43	0.004	0.007	0.023	0.058	0.149	0.293	0.339	0.268
48–53	0.002	0.001	0.004	0.007	0.017	0.055	0.146	0.165
60+	0.001	0.001	0.008	0.007	0.009	0.050	0.134	0.357

Figure 2 illustrates the overall system pipeline. A face is first detected and aligned to a canonical pose using a 2D alignment tool [2]. The aligned face is resized to 256×256 and fed into the CNN. The CNN outputs a probability distribution over age groups or gender. During oversampling, multiple crops of the face are evaluated and their softmax outputs averaged. The final prediction is the class with highest probability.

VII. DISCUSSION AND CONCLUSION

Our CNN-based approach achieves state-of-the-art performance on the Adience benchmark [1]. We repli- cated the results of Levi and Hassner, with gender accuracy around 86.8% and age exact accuracy around 50.7%. This significantly surpasses earlier methods [2]. The confusion matrix (Table IV) confirms that most errors are off by one age group, highlighting the intrinsic ambiguity in human aging.

Training the CNN from scratch requires a few hours of GPU time, and oversampling increases inference

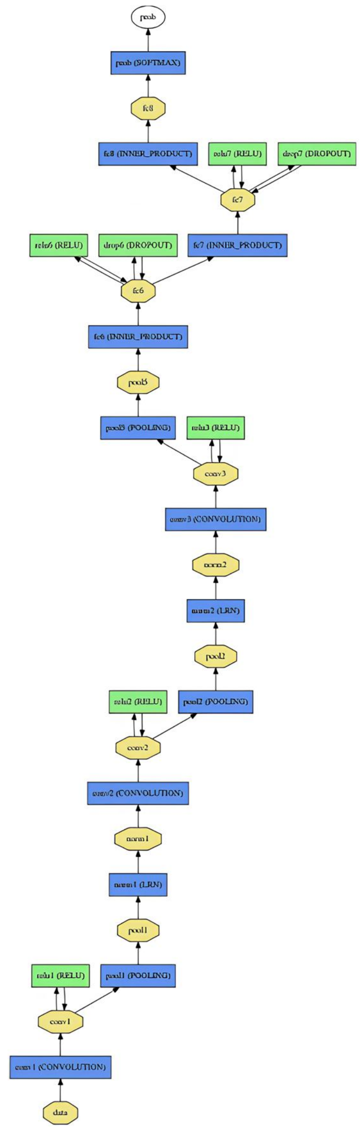

Fig. 2. System pipeline: face detection and alignment followed by CNN-based age and gender classification.

cost, but these are reasonable given the accuracy gains. Future work could involve training deeper or pretrained networks with larger face datasets, or improving pre- processing (e.g., advanced face frontalization) to reduce errors. Regression-based age estimation and multi-task learning are also promising directions.

In summary, this report provides a detailed technical account of CNN-based age and gender classification on unconstrained images. We cover the Adience dataset, the CNN model architecture, training and augmentation strategies, and evaluation results. The strong perfor- mance on Adience demonstrates the power of CNNs for this task and provides a solid baseline for future research.

REFERENCES

[1] G. Levi and T. Hassner, "Age and Gender Classification using Convolutional Neural Networks," in Proc. IEEE CVPR Workshops, 2015.
[2] E. Eidinger, R. Enbar, and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," IEEE Trans. Inf. Forensics Security, vol.9, no.12, pp.2170–2179, 2014.
[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classifi- cation with Deep Convolutional Neural Networks," in Adv. Neural Inf. Process. Syst., 2012.
[4] G. E. Hinton et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," J. Mach. Learning Res., vol.15, pp.1929–1958, 2014.