# AN ENSEMBLE CNN2ELM FOR AGE ESTIMATION

Therefore, age estimation should not only consider race and gender as factors but also yield a novel structure in which differentiable features can be extracted. Motivated by the analysis above, we propose a novel age estimation architecture called CNN2ELM, which is used to estimate age based on face images. CNN2ELM includes three convolutional neural network structures and two extrame learning machine (ELM) structures. Our proposed networks are pretrained on the ImageNet dataset and then fine-tuned on the IMDB-WIKI dataset [22] before being fine-tuned on MORPH-II, Adience benchmark, and ChaLearn Looking at People 2016 (LAP-2016) datasets. During the validation and test stages, the three networks are used to extract features from the same image, and a discriminative and robust feature set is achieved by fusing these features. Based on the fusion features, ELM classifies these features into one of the age groups. It not only sufficiently exploits the CNN but also utilizes the outstanding classification and regression properties of the ELM. The major contributions of this paper are as follows:

* We propose an ensemble CNN2ELM approach for performing age estimation. The structure combines the CNN with the ELM to perform age estimation in a hierarchical fashion, in which the CNN is used to extract features, while the ELM predicts age based on the image of a person age. The system full utilizes the advantages of the CNN and ELM.
* The highlights of our CNN2ELM structure are feature enhancement and age grouping. Our system comprises three CNNs and two ELMs (i.e., Race-Net + AgeNet + Gender-Net + ELM classifier + ELM regressor (RAGN)”). **TABLE 1**
* Finally, we present the process of integrating the synergyof the hybrid structure in detail, including the design of the layers in the CNN. By a comprehensively considering the factors mentioned above and after many experiments, we set the size of the input feature map as 227 × 227, which is showed in **FIGURE 1** .

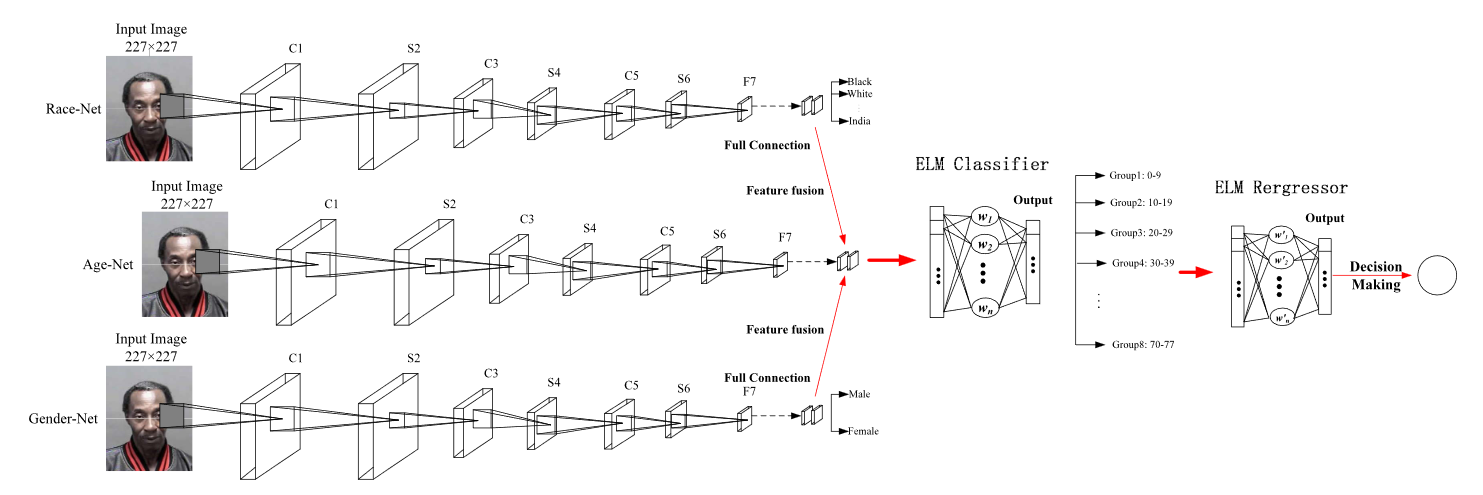


FIGURE 1 Full schematic diagram of our network architecture.

# 

# RELATED WORK

## Hybrid Neural Network

System CNN has been successfully applied to various fields, and image recognition, in particular, is a hot area of research. The approach reduces the number of parameters in the network and reduces the training time. Based on the above results, we decided to integrate CNN with other classifiers to improve the classification accuracy. In Sections IV, we will present the ensemble CNN2ELM in detail and show that it has better performance compared with other methods for processing the same tasks.

## Fusion Methods

Fusion is a popular technology in biometrics, and it is most commonly used to fuse decisions or features in a hierarchical learning system. One of the most successful examples is the ensemble system [20]. This system proves that several classifiers with similar training characteristics have different generalization performances, and fusing these classifiers may or may not lead to a better performance than that of the best classifier in this ensemble system while reducing the global risk of making a poor decision.

## Age Grouping

In recent years, facial age grouping has been widely used for age estimation and many methods have been proposed. Kwon and da Vitoria Lobo [5] first introduced the age grouping method for age estimation, which starts by categorizing images into three age groups: infants, youth, and seniors. During the experiments, 47 images were used to verify the performance of the proposed method, and the classification accuracy for the infant group was below 68%.

For Gender-Net, the selected dataset and settings of the network and parameters are the same as for Race-Net. **FIGURE 2** shows the process of Gender-Net.

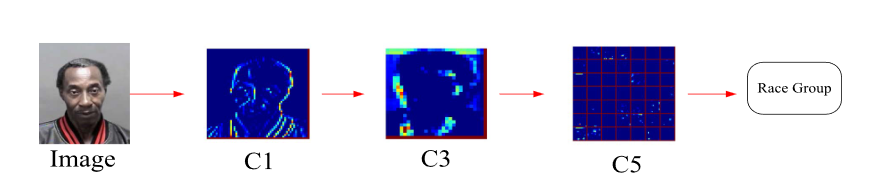


FIGURE 2 The process of Race-Net. The color indicates the output of convolutional layers.

Next, the outputs of the convolutions add an additive bias and an element-wise nonlinear activation function is then applied to the front results. We have (1)

|  |  |
| --- | --- |
| , | (1) |

where represents the bias of this feature map, and δ indexes over the set of the feature maps in the (i −1)th layer, which are connected to this convolutional layer. denotes the value at the position ( p, q) of the kernel, which is connected to the kth feature map.

# 

# PRELIMINARY INFORMATION

The purpose of our feature vector fusion is to obtain an enhanced feature vector that is beneficial to the ELM classifiers. **FIGURE 3** shows the process of feature fusion. Because classifiers learn more correlative and useful information, they can achieve a better generalization performance.

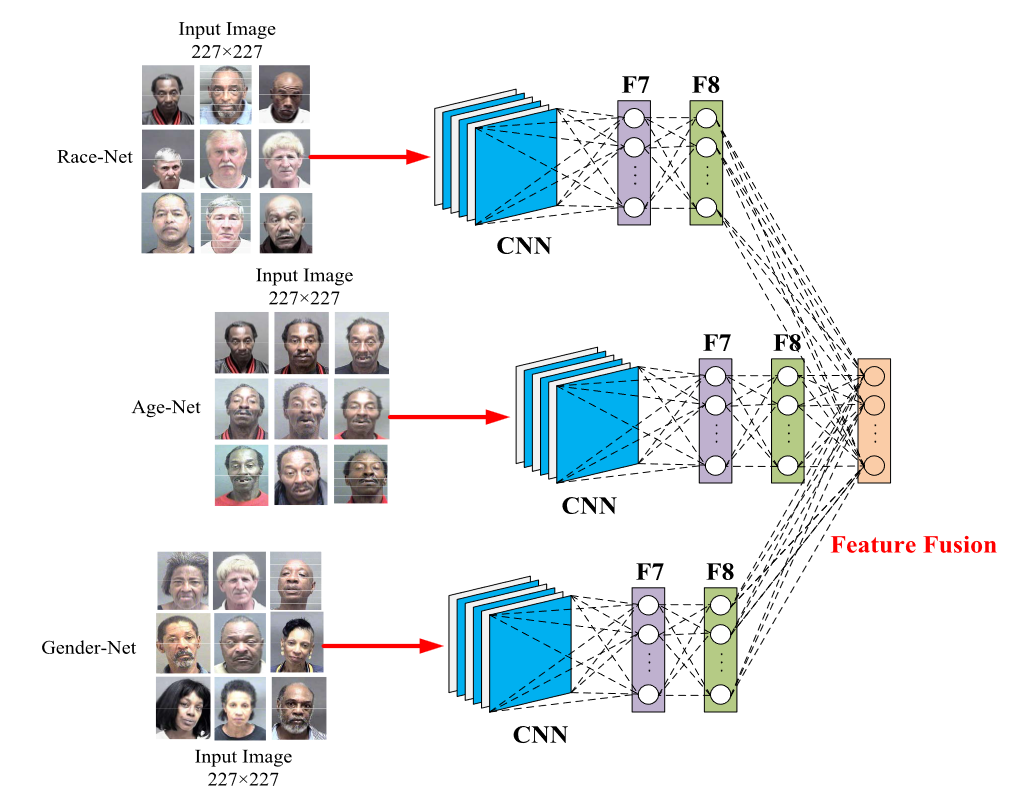


FIGURE 3  The process of feature fusion. During the validation and test stage, Race-Net, Age-Net, and Gender-Net are used to extract features from the same image. The Design of Our Ensemble Structure:

1. Convolutional Layer;
2. Contrast Normalization Layer;
3. Max Pooling Layer.

To achieve the target, two normalization operations, i.e., subtractive and divisive, are performed. ηmnk denotes the value of an unit at position (m, n) in the kth feature map. We have (2)

|  |  |
| --- | --- |
| , | (2) |

where is a normalized Gaussian filter with the size of 7 × 7 at the first stage and 5 × 5 at the second stage. zmnk not only represents the input of the divisive normalization operations but also denotes the output of the subtractive normalization operations. Equation (3) expresses the operator of the divisive normalization:

|  |  |
| --- | --- |
| , | (3) |

where M(m, n) = (4)

|  |  |
| --- | --- |
|  | (4) |

|  |  |  |  |
| --- | --- | --- | --- |
| TABLE 1  The Data Set Selected From MORPH-II For Our Experiment | | | |
|  | Female | Male | Female and Male |
| Black | 5,757 | 36,809 | 42,560 |
| White | 2,601 | 7,999 | 10,600 |
| Black and White | 8,358 | 44,802 | 53,160 |