AGE ESTIMATION AND GENDERRECOGNITION USING CNN

Tejonidhi S[1*], Sudha BG[2], Narayana Darapaneni[3], Anwesh Reddy Paduri[4]

[1]AIML, PES University, Hosur Rd, Konappana Agrahara, Bangalore,560100, Karnataka, India.
[2]Faculty & Senior Manager, Data Science and AI, Great Learning.
[3]Director - AIML, PES University, Hosur Rd, Konappana Agrahara,Bangalore, 560100, Karnataka, India.
[4]Senior Data Scientist, PES University, Hosur Rd, Konappana Agrahara, Bangalore, 560100, Karnataka, India.

*Corresponding author(s). E-mail(s): tejonidhi1997@gmail.com; Contributing authors: sudha.m@gmail.com; darapaneni@gmail.com; anwesh@greatlearning.in;

Abstract

Computer vision facial image analysis presents significant difficulties when trying to accurately estimate age and identify gender. The goal is to address this challenge and our project uses Convolutional Neural Networks (CNN) to predict age and gender from facial images. Our model is trained and evaluated using the UTK dataset, a large collection of facial images, associated with age and gender markers. Carefully designed and trained, our CNN model detects subtle facial features that indicate age and gender. Early observations show very promising accuracy rates, showing the potential of CNNs in this field. The results of this project have important value for the growth and development of the industry.

Keywords: Computer vision, UTK Face, VGG16, Convolution Neural Network, Age, Gender.

1 Introduction

Facial attribute analysis, including gender detection and age estimation, has received considerable attention in computer vision research due to its wide range of applications in various fields. Recent advances in deep learning techniques have shown significant success in solving these tasks using large datasets to train highly accurate models. In this paper, we present in detail deep learning approaches for gender detection and age estimation in face images, focusing on the use of the UTK dataset. Sexual orientation acknowledgment from facial pictures includes classifying people into male or female categories based on facial highlights, whereas age estimation points to anticipate the age of people delineated within the pictures. These assignments are inalienably challenging due to varieties in facial expressions, lighting conditions, and age-related changes in appearance. In any case, profound learning models, especially convolutional neural frameworks (CNNs), have outlined transcendent execution in learning solid representations from complex visual data. Our proposed model is built upon a convolutional neural network (CNN) architecture, trained end-to-end to predict gender labels and estimate age from facial images in a single forward pass. By exploiting common features, our model learns to capture both sex-specific and age- related facial features, which facilitates mutual information transfer between the two tasks. The basis of our research is UTK, which is famous for its large collection of facial images with age and gender characteristics. Through extensive tests and evaluations, we demonstrate the effectiveness of our integrated model in accurately predicting gen- der and age estimation across a variety of demographic groups and imaging conditions in our dataset.

2 Literature Survey

The paper[2] presents a novel strategy utilizing Independent Component Analysis (ICA) for gender classification through facial images. The researchers started by normalizing facial images to account for variations in geometry and illumination, primarily based on eye location. The normalized images were then processed to represent them in a lower-dimensional space using ICA. This representation was further used for gender identification, employing various classification algorithms. The experiments utilized the FERET facial dataset, which consists of frontal shots of individuals. Using their approach, the researchers achieved a classification accuracy of up to 95.67% using Support Vector Machine (SVM) in ICA mode. This approach offers a promising direction for gender classification using facial images, potentially paving the way for more advanced applications in biometrics and human-computer interaction. Another study [3] delved into the challenges of face analysis from unconstrained images. They highlight the difficulties caused by differences in resolution, deformation, and occlusion. To address these challenges, the authors introduced a new mechanism of attentional bias. This mechanism is designed to identify the most informative and reliable parts of the face, improving age and gender classification. The proposed model first works by reducing the face image. It then uses a full learning framework to extract the most characteristic locations from the original high-resolution image. The researchers validated their approach using standard benchmarks such as Adience, Images of Groups, and MORPH II. Their results show that the inclusion of attentional mechanisms improves the robustness and accuracy of convolutional neural networks (CNN) in age and gender detection tasks. One way to detect gender was using features like cheeks, hair, and skin [4], achieving an accuracy of 78%, but combined with the nose vector, it reaches 90%. Another method is principal component analysis (PCA) and linear discriminant analysis (LDA) [5][6]. There is another generalized dataset, LFW, where the CNN ensemble method was used to detect gender when highly significant [7]. Another algorithm dealing with face detection, orientation localization, pose estimation, and gender detection was given by [8], which combines CNN intermediate layers. Another approach is fine-tuning the pre-trained VGG16 model [8] for automatic gender detection. Using Deep Convolutional Neural Network (DCNN) for input image prediction and feature selection and finally using Support Vector Machine (SVM) for age and gender classification [9]. The authors use the FG-NET Aging Database [9] and the MORPH database [10][11] in the experiment.

3 Materials and Methods

3.1 Dataset Collection and Preparation

The quality and differences of the dataset play a crucial role in the development and evaluation of deep learning models for gender recognition and age estimation tasks. Here we detail the process of collecting and preparing the UTK dataset with specific techniques including the selection of age groups and random blurring which serve as the foundation of our research.

To enhance the model's ability to generalize across various conditions, we incorporated additional augmentation techniques such as random rotations, horizontal flipping, and brightness adjustments. These augmentations aimed to simulate real-world variability and improve the robustness of the model. Additionally, a portion of the dataset was reserved as a test set, which was not seen by the model during training, ensuring an unbiased evaluation of the model's performance.

The dataset was also analyzed for any class imbalances, and techniques such as oversampling and class weighting were considered to address any discrepancies. This ensured that the model received a balanced representation of each class during training. The preprocessing steps included normalizing the pixel values to a range of [0, 1], which facilitated faster convergence during training.

By meticulously preparing the dataset, we aimed to create a solid foundation for training a deep learning model capable of accurately predicting age and gender from facial images.

3.2 Dataset Description

The UTK dataset is a freely accessible benchmark dataset comprising facial images annotated with age, gender, and ethnicity labels. It consists of a diverse collection of over 20,000 images capturing individuals of different ages, genders, and ethnic backgrounds. For our research, we focused on creating a subset where each age group up to 60 years old is represented by 100 images, resulting in a balanced distribution across age categories. In addition to the age, gender, and ethnicity labels, the dataset provides images in various resolutions and lighting conditions, which adds to the complexity and robustness of the training process. The UTK dataset includes individuals from a wide range of ethnicities, making it an excellent resource for developing models that generalize well across different demographic groups. The diversity in age groups, from infants to the elderly, allows for comprehensive training and evaluation of the model across the entire human lifespan. The inclusion of multiple ethnic backgrounds helps in assessing the model's performance in a globally diverse context. Such a well-rounded dataset is critical for developing models that are not biased towards any specific group and can perform reliably in real-world applications.Overall, the UTK dataset serves as a robust foundation for our research, providing the necessary diversity and complexity to train a high-performing age and gender prediction model.

3.3 Data Preprocessing

Prior to model training, the collected images underwent preprocessing steps to enhance compatibility with deep learning frameworks and to simulate age progression. A notable preprocessing step involved randomly darkening 30% of the images within each age group to simulate the effects of aging. This augmentation technique aimed to introduce variability in facial appearance and improve the model's robustness to age-related changes.

3.4 Data Visualization

The bar chart displayed in Figure 1 outlines the distribution of a population sample by age and gender. Each bar represents the count of individuals within a specific age range, with blue bars indicating male and pink bars indicating female individuals. The x-axis indicates the age groups, while the y-axis measures the number of individualswithin each group.

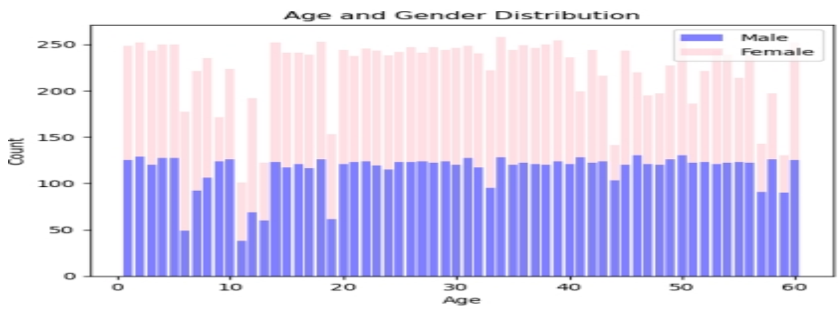


Fig. 1 Age Distribution Chart

3.5 Model Architecture

Our deep learning model architecture is constructed using Tensor Flow and Keras. The architecture is designed to effectively capture gender-specific and age-related facial features from input images. It comprises several convolutional layers followed by batch normalization and max-pooling layers to extract hierarchical representations of facial attributes. The input layer is characterized by a shape of (128, 128, 1), representing grayscale images with a resolution of 128x128 pixels. This input is fed into a series of convolutional layers, each with increasing numbers of channels and kernel sizes, to learn spatial features at different scales. Batch normalization layers are incorporated after each convolutional layer to stabilize and accelerate the training process. Following the convolutional layers, max-pooling layers with a pool size of (2, 2) are applied to down sample the feature maps, reducing their spatial dimensions while retaining important information. This helps in creating translation-invariant representations of facial features. After the convolutional blocks, a global average pooling layer is used to aggregate spatial information across the feature maps, resulting in a compact representation of the input images. This aggregated representation is then fed into separate branches for gender prediction and age estimation. The gender prediction branch consists of fully connected dense layers with ReLU activation functions. Dropout regularization is applied to prevent overfitting, and the output layer utilizes a sigmoid activation function to output the probability of being male (0 to 1).Similarly, the age estimation branch comprises fully connected dense layers with ReLU activation functions. Dropout regularization is again applied, and the outputlayer uses a linear activation function to output the estimated age. The model is instantiated using the Model class from Keras, with the input and output layers specified accordingly. Additionally, another pre-trained model, VGG16, with weights from ImageNet and an input shape of (220, 220, 3), was fine-tuned. To prevent overfitting, the first 15 layers were frozen, and two dense layers with 256 units each and ReLU activation functions were added. To enhance the diversity and generalization capability of our model, data augmentation techniques such as rotation, flipping, and scaling were applied. Subsequently, we integrated two distinct output layers to our model: one for predicting age using a linear activation function and another for predicting gender utilizing a sigmoid activation function.

3.6 Training Procedure

The training procedure involves optimizing the model parameters to minimize the loss functions for gender prediction and age estimation. We utilize the Adam optimizerwith a learning rate of 0.0001 to efficiently update the model weights. The binary cross-entropy loss is employed for gender prediction, while the mean squared error loss function is used for age estimation. During training, the model's performance is monitored using evaluation metrics such as accuracy for gender classification and mean absolute error (MAE) for age estimation. These metrics provide insights into the model's performance on bothclassification and regression tasks. To prevent overfitting and improve generalization, several callbacks are utilized, these include Early Stopping, ModelCheckpoint and ReduceLROnPlateau.

3.7 Evaluation Metrics

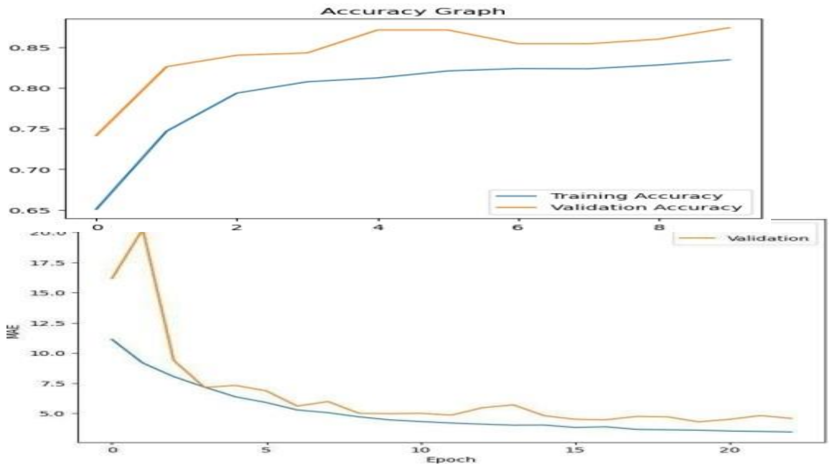
The performance of the trained model is evaluated using various metrics, including accuracy, precision, recall, F1-score, MAE, and RMSE. These metrics are computedon both the training and validation datasets to assess model generalization and overfit- ting. The model's ability to correctly classify individuals into male or female categories and predict their ages accurately is analyzed and discussed.

3.8 Experimental Setup

Experiments are conducted on the UTK dataset to assess the performance of the proposed model. The dataset is split into training, validation, and testing sets following a standard ratio. Hyper parameters such as learning rate, batch size, and dropout rate are tunedto optimize model performance. Cross-validation is also used to evaluate model robustness and stability.

4 Results

4.1 Age Prediction Performance



Mean Average Error (MAE): Figure 2 outlines the model’s performance in predicting age, measured by MAE over epochs for both training and validation sets. The training MAE (blue line) and validation MAE (orange line) both show a pronounced decrease in the initial few epochs, indicating rapid learning by the model. The MAE values level off after the 5th epoch, with the validation MAE showing slight variability but remaining relatively stable, which implies that the model is generalizing well without overfitting to the training data.

Fig. 2 Age mean absolute Error

Age Loss: As shown in Figure 3, the loss associated with age prediction decreases sharply at the outset for both the training (blue line) and validation (orange line) datasets, corroborating the findings from the MAE analysis. The convergence of training and validation loss values suggests that the model is learning to predict age with a degree of certainty that is consistent across both datasets, reinforcing the model's generalization ability.

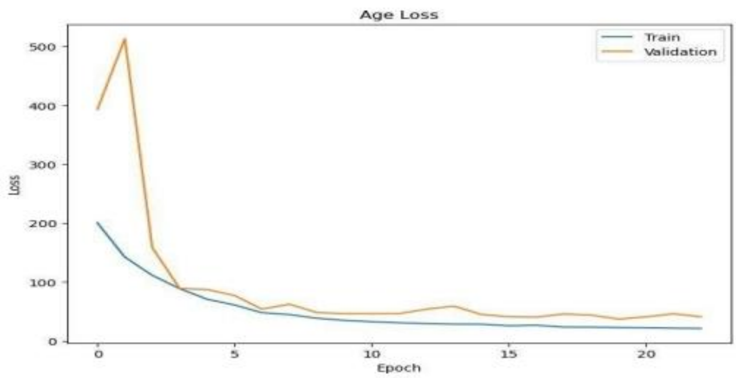


Fig.3 Age Loss Graph

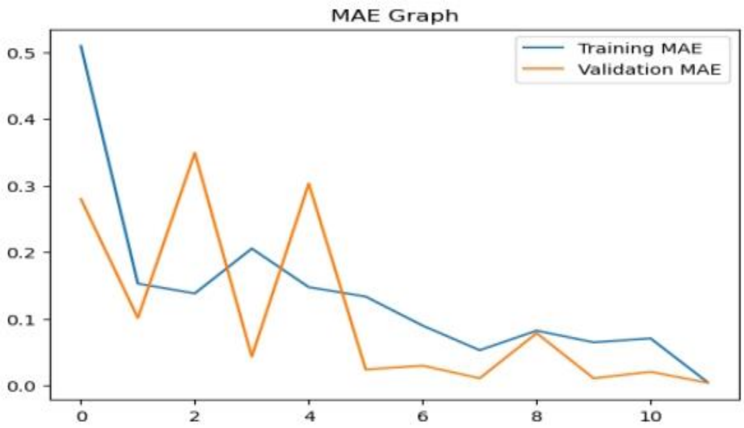


Fig.4 Age MAE Graph of VGG16

The graph shows the Mean Absolute Error(MAE) for both training and validation sets over approximately 11 epochs of a machine learning model's training process. The MAE for the training data(blue line) starts high and shows a sharp decrease, then fluctuates while generally trending downward. The validation MAE(orange line) mirrors the downward trend with less fluctuation. Both metrics decreasing over time suggests the model is improving and generalizing well to unseen data, with no obvious signs of overfitting. However, the volatility in the training MAE suggests potential instability in the training process, which might be improved by adjusting the model's hyperparameters. Model prediction of Age is of by 4 years.

Fig.5 Comparison of Age and Gender Prediction

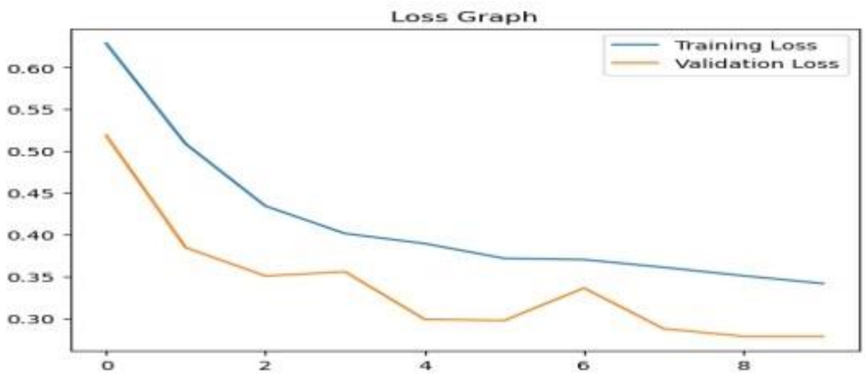


Fig. 6 VGG16 Loss Graph

Gender Loss of the VGG16 model as we can see in the Fig6 both training and validation accuracy are very high and close to each other suggests that the model is performing very well on both the data it was trained on and the unseen data. But it also shows that there are chances of overfitting as the model seems to perform too well.

Fig. 7 Real time result of VGG16

From **Fig7** we can observe that the real time result of the image validated the gender was correctly classified and the Age was off by 4 years.

Fig. 8 Real time result of custom model

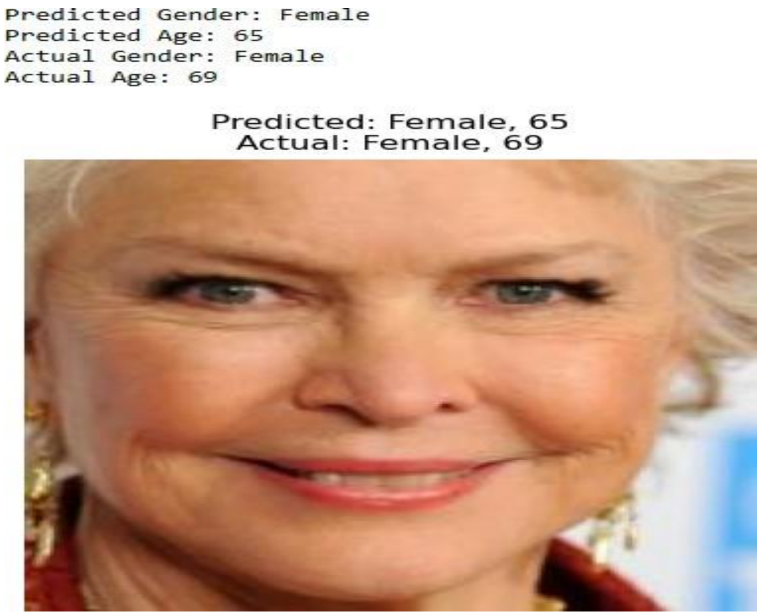


Fig. 9 Real time result of VGG16 over age 60

From the **Fig8** we can observe that the real time result of the image validated the gender was correctly classified and the Age was off by 3 years and w.r.t **Fig9** we cansee for age more than 60 the age estimation is off by 4 years.

Fig. 10 Real time result of custom model over age 60

From the **Fig10** we can observe that the gender classification is correct and the age estimation is off by 9 years

5 Conclusion

This research aimed to develop a neural network model that simultaneously classifies gender and predicts age, employing a multi-task learning framework. The results obtained from extensive training and validation phases have provided several insights into the model's learning capabilities and generalization performance. The gender classification component of the model achieved a good validation accuracy, peaking at 86.67%. This level of accuracy underscores the model's robustness in gender classification tasks, demonstrating the efficacy of the chosen network architecture and training regimen. The model's accuracy in gender prediction improved

steadily over the epochs, indicating that the model was learning effectively without being prone to significant overfitting, as the loss for gender classification remained low and stable. Age prediction, while showing a decrease in training and validation MAE, suggested a more complex pattern. The training MAE consistently decreased, showing the model's increasing precision in predicting age. The validation MAE, after reaching its best value at epoch 20, showed a slight increase in subsequent epochs. This difference between training and validation performance possibly suggests the beginning of overfitting or the difficulty in age prediction from the provided dataset, which might contain varied patterns that are not fully captured by the model. The loss graphs illustrate the model's initial rapid learning phase, followed by stability, which is typical in machine learning and signifies a saturation point where the model starts to yield decreasing returns on learning from the training data. The stability in validation loss and MAE also indicates that the model reached a level of generalization beyond which further training does not significantly improve performance on the validation dataset. Many steps were taken to fix overfitting and enhance model performance, such as model checkpointing, which effectively captured the best state of the model for both tasks. The results from this study contribute to the growing body of knowledge in multi-task learning, illustrating that a single model can effectively learn and predict multiple outputs simultaneously. While the results are promising, they also highlight the challenges in age prediction tasks, which could be addressed in future work. Potential directions include exploring more complex model architectures, incorporating a larger and more diverse dataset, and implementing advanced regularization techniquesto enhance the model's predictive power and generalization.
There are some drawbacks to this model as we move further to test any image more than 40 years of age; there is a rather huge gap in age prediction, but with respect to ages up to 40, it would be off by a maximum of 4 years. The VGG16 model has a similar performance in terms of age for comparison, and for gender classification, it identifies correctly. The VGG16 is 87% accurate for gender classification, and for age estimation, it is off by 4.4 years. In conclusion, our research provides a solid foundation for future explorations into multi-output demographic prediction models. The successes and challenges encountered point to the importance of continuous innovation in model development and training techniques within the field of machine learning. However, the model's performance exhibited certain limitations, particularly in age prediction for individuals older than 40 years. The age prediction accuracy tends to decrease with increasing age, indicating a need for further refinement in handling age-related facial changes. The VGG16 model, used for comparison, showed similar performance trends, highlighting the consistent challenges in this task across different architectures. Despite these challenges, the research underscores the potential of deep learning models in complex tasks like age and gender prediction. Our findings highlight the importance of continuous refinement in model training and the integration of more diverse datasets to improve generalization. By addressing these limitations and exploring advanced techniques, future work can build upon the solid foundation laid by this study, pushing the boundaries of what is possible in age and gender prediction from facial images.

References

[1] Hassan, K. R., & Ali, I. H. (2020). Age and Gender Classification using Multiple Convolutional Neural Network. *IOP Conference Series: Materials Science and Engineering*, 928(032039). doi:10.1088/1757-899X/928/3/032039.

[2] Jain, A., Huang, J., & Fang, S. (2005, July). Gender identification using frontal facial images. In 2005 IEEE International Conference on Multimedia and Expo (pp. 4-pp). IEEE.

[3] Gunay, A., & Nabiyev, V. V. (2008, October). Automatic age classification with LBP. In 2008 23rd international symposium on computer and information sciences(pp. 1-4). IEEE.

[4] Bayana, M., & Viriri, S. (2018, March). Gender classification based on fusion

of facial components features. In 2018 Conference on Information Communications Technology and Society (ICTAS) (pp. 1-5). IEEE.

[5] Hapani, S., Prabhu, N., Parakhiya, N., Paghdal, M. (2018). Automated Atten-dance System Using Image Processing. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, pp. 1-5. DOI: 10.1109/ICCUBEA.2018.8697824.

[6] Jin, K., Xie, X., Wang, F., Han, X., Shi, G. (2019). Human Identifica-tion Recognition in Surveillance Videos. In 2019 IEEE International Confer- ence on Multimedia & Expo Workshops (ICMEW). IEEE, pp. 162-167. DOI: 10.1109/ICMEW.2019.00-93.

[7] Antipov, G., Berrani, S. A., & Dugelay, J. L. (2016). Minimalistic CNN-based ensemble model for gender prediction from face images. *Pattern recognition letters*, 70, 59-65.

[8] Dhomne, A., Kumar, R., & Bhan, V. (2018). Gender recognition through face using deep learning. *Procedia computer science*, 132, 2-10.

[9] The FG-NET Aging Database, http://sting.cycollege.ac.cy/ alani-tis/fgnetaging/index.htm, 2002.

[10] Ricanek, K., & Tesafaye, T. (2006). MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. In Proceedings of the Seventh International Conference on Automatic Face and Gesture Recognition, pp. 341-345.

[11] Geng, X., Zhou, Z. H., & Smith-Miles, K. (2007). Automatic Age Estimation Based on Facial Aging Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12), 2234-2240. doi: 10.1109/TPAMI.2007.70733.