Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation

Philip Smith
Computer Science
UNC Wilmington
North Carolina, United States
ps1994@uncw.edu

Cuixian Chen
Mathematics and Statistics
UNC Wilmington
North Carolina, United States
chenc@uncw.edu

***Abstract*—In this project, competition-winning deep neural networks with pretrained weights are used for image-based gender recognition and age estimation. Transfer learning is explored using both VGG19 and VGGFace pretrained models by testing the effects of changes in various design schemes and training parameters in order to improve prediction accuracy. Training techniques such as input standardization, data augmentation, and label distribution age encoding are compared. Finally, a hierarchy of deep CNNs is tested that first classifies subjects by gender, and then uses separate male and female age models to predict age. A gender recognition accuracy of 98.7% and an MAE of 4.1 years is achieved. This paper shows that, with proper training techniques, good results can be obtained by retasking existing convolutional filters towards a new purpose.**

***Index Terms*—transfer learning, deep learning, convolutional neural network, age estimation, gender classification**

I. Introduction

Deep learning has developed rapidly recently due to the availability and large scale of labeled data, and high performance computing. Deep learning is currently one of the most popular machine learning techniques in Artificial Intelligence (AI) [1], [2], [3], [4], [5]. In computer vision, many classical techniques of feature extraction and subspace learning have been eclipsed due to the recent good performance in deep learning. Traditionally, good results have been obtained using combinations of classifiers, regressors, hand-crafted features, facial landmarks, and dimensionality reduction [6] [7] [8]. Neural networks, however, have taken the scene with their ability to learn and memorize features, and keep improving in accuracy as more data are observed. Thus Deep Neural Networks (DNNs) have far surpassed traditional classification and regression techniques, and have even surpassed human performance on a number of well-known benchmarks [9] [10].

II. Related Works

In a modern interconnected information society, it is critical to identify or verify individuals accurately at real-time. Due to its significant role in human computer interaction (HCI), internet access control, and security control and surveillance, face-based demographical research has attracted great attention in both research communities and industries [11]. MORPH-II [12] has been the subject of many studies concerning age and gender estimation. As such, it is a good way to compare the efficacy of different techniques. Han et al. [9] gauges human age estimation by crowd-sourcing estimates on two popular face-image databases. They found estimates on the FG-NET dataset to be off by an average of 4.7 years. They mention that that number might be low because it is easy to guess the ages of babies and children without much variation in predictions. In fact, the average age error on FG-NET subjects older than 15 is 7.4 years, which is similar to the human error of 7.2 years that was calculated on the PSCO dataset. In the same study, Han et al. use a hierarchy of support vector machines (SVMs) and biologically-inspired features (BIFs) to obtain an average age estimation error of 4.2 years on the MORPH-II database.

Deep learning is promising to allow for the full utilization of large datasets in order to solve machine learning problems. Amongst the different types of deep learning architectures, convolutional neural networks (CNN) have been proven to be very effective for human demographics estimation due to their proficiency at extracting precise details from images. Such studies include age estimation [13], [14], [15] and gender classification [16], [17], [18] . Niu et al. [19] obtain an error of 3.28 years using ordinal regression CNNs and random splits of the MORPH-II dataset where 80% of the images are used for training and 20% are used for testing.

Rothe et al. [13] considered deep CNNs for age classification problems. The VGG-16 architecture and IMDB-WIKI dataset are employed in this study. With a random split of 80% for training and 20% for testing on MORPH-II, it achieves a MAE of 2.68 with additional fine-tuning on IMDB-WIKI dataset before fine-tuning on MORPH-II dataset. Later, Antipov et al. [20] extend the work from [13] and consider the problems of selection of optimal CNN architecture and training strategies. They conclude that Label Distribution Age Encoding (LDAE) [21] is an optimal way for the target encoding to train a CNN for an age estimation task. It is showed that face recognition pretraining is more effective for deep gender and age CNNs comparing to general task pretraining. Following the subsetting scheme in [22] for MORPH-II, it achieves a MAE of 2.99 years with VGG-16 pretrained CNN for facial recognition, and a gender classification accuracy of 99.3% with ResNet-50 pretrained CNN for facial recognition. Their model also won the ChaLearn Apparent age estimation

challenge in 2016 [16].

In this paper, transfer learning is employed to tackle the problem of recognizing a person's age and gender from an image using deep CNNs. A variety of network designs and training techniques are explored. We consider dynamic LDAE, which outperforms the static LDAE considered in [20]. A gender-specified hierarchical age model is proposed in this study. Experimental results demonstrate its effectiveness over the general age model.

III. Transfer Learning

Because of the vastness and complexity of deep neural network architecture, designing and testing models is expensive and time-consuming. When approaching an AI problem, quick results can be obtained by utilizing a technique known as *transfer learning*. In transfer learning, the weights and convolutional filters that are proficient at one task, can be reused for a different task requiring only a small amount of retraining. This involves using a network architecture with preloaded weights, modifying it slightly, and then retraining part or all of the model to output predictions for the new task. The filters learned by one task, such as classifying animals, are used to extract features from images that can then be interpreted by the retrained portion of the neural network in order to perform its new task. In this paper, the deep convolutional neural network known as VGG [23] is used to study transfer learning using two different types of pretraining.

A. VGG19

VGG19 [23] is a DNN architecture developed by Karen Simonyan and Andrew Zisserman of the Visual Geometry Group at Oxford. The "19" in the name refers to the number of weight layers in the network. VGG16 was considered to be more successful in the ImageNet competition in 2014 and tied with GoogLeNet, however, the extra depth of VGG19 was leveraged to achieve better results than VGG16 in some instances. The original VGG architectures consist of five stacks of convolutional layers, each followed by max pooling layers. The top layers are the same across all VGG designs and consist of two fully-connected layers, each of size 4096 with 50% dropout, and a fully-connected softmax layer of size 1000.

B. VGGFace

Shortly after the release of the VGG architectures, the Visual Geometry Group published another paper called "Deep Face Recognition" [24]. In this paper, VGG16 is trained from scratch for facial recognition using a dataset of 2.6 million face images. Prior studies have shown that transfer learning using neural networks with facial recognition pretraining can produce highly effective results for gender recognition and age estimation [16]. Since facial recognition neural networks have already been trained to distinguish human features, the features that they extract may be more useful for determining age and gender from a photo than the features extracted by a more general neural network. In this study, VGGFace, VGG16 with facial recognition weights, is also examined for its proficiency at age and gender classification.

IV. The Datasets

Machine learning models rely on the quality of data that feed them. Mislabeled data and excessive noise can cause models to start learning the wrong things. In deep learning, large and accurate datasets are essential to obtaining good performance. In this study, the MORPH-II dataset is used to train and test models.

A. MORPH-II

MORPH-II [12] is a good candidate for gender and age or other face image studies for a few reasons. The images captured are of the subjects' heads and most are positioned in front of a gray background – which helps reduce background noise. Age labels and other information are provided about the subjects such as race, gender, and a unique identifier. A visual inspection of the images, however, reveals a few noisy variation. The subjects' heads are tilted in different directions, and may be of varying distance from the camera. Pixelations are apparent in most images, and some images have vastly different tint. The dataset consists of 55,134 images with subject ages ranging from 16 to 77 years old. 84.6% of the dataset is male, and 77.22% of the dataset is black. As seen in table I, few images exist of subjects 50 years of age and older. Because of this, a subsetting strategy has been adopted by the academic community from the works of Guo and Mu [22]. They propose to divide the dataset into three subsets. The first two, S_1 and S_2, consist of only blacks and whites, and have a 3:1 male to female ratio. S_3 contains all of the remaining images.

TABLE I
Age by Gender in MORPH-II

	<20	20-29	30-39	40-49	50+	Total
Male	6649	14009	12436	10082	3468	46644
Female	836	2305	2924	1978	447	8490
Total	7485	16314	15360	12060	3915	55134

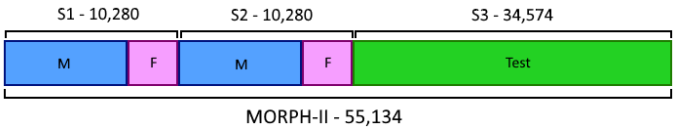

Fig. 1. A depiction of subsetting scheme on a cleaned version of MORPH-II, following the subsetting proposed by Guo and Mu [22].

B. MORPH-II Cleaned

A survey of the MORPH-II dataset revealed several inconsistencies. Some subjects had different dates of birth. Others had multiple race labels or both gender labels. In order to combat the effects of "dirty data", such inconsistency in age, gender and race has been manually identified and cleaned up for MORPH-II dataset. More details can be found in [25]. Hereafter, the MORPH-II cleaned data are used in this project. Following the subsetting scheme proposed in [22],

our subsetting scheme is shown in figure 1. Sets 1 and 2 both contain 10,280 images, and set 3 contains 34,344 images.

C. MORPH-II Equalized

Preprocessed MORPH-II dataset are also considered in our preliminary study for performance evaluation. In this case, MORPH-II images are first cropped to fit the subjects' faces. During the process, the images are rotated such that the subjects' eyes are aligned. Images are grayscaled, and the lighting of the images is equalized. This dataset, known as MORPH-II equalized, is used in the early testing stages of our preliminary study due to its small input vector size [26]. The full-sized images are either 200x240 or 400x480 having input vector sizes of 144,000 and 576,000 respectively, but the equalized images only produce 4,200 data points.

V. Training Parameters

To compare the effects of changes in transfer learning techniques, all training parameters are kept consistent unless otherwise specified. For MORPH-II, all images are scaled down to 200x240. All input is standardized before being fed into the network. The batch size is set to 50, and models are trained for 60 epochs. The original dropout rate of 0.5 is retained, and the ReLU activation function is used in all weight layers. The Adadelta optimizer is used with its default values. Gender models use the binary cross entropy loss function, and age models use mean absolute error (MAE). Results for age estimation are reported as an MAE, which is defined as:

$$\mathrm{M}AE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}|. \tag{1}$$

As such, MAE is the average of the absolute differences between the predicted age and the subject's actual age. For gender, the results are reported as an accuracy – the number of correct predictions over the size of the test set.

S_1 is used to train the models. During the parameter training process, models are supplied with a validation set of 500 random samples from S_3. To show the performance of models as data are added, the training set is split into several sets that are trained upon serially. This also helps avoid the issues that arise from using too much computer memory. The model parameters with the lowest loss on the validation set sample is saved and then fully validated on $S_2 \cup S_3$, a set of 44,624 images.

VI. Top-Layer Retraining

A common practice in transfer learning is to remove the top layers of a DNN, and then replace them with a different top. In VGG19, the top of the network is responsible for interpreting the output of the many underlying convolutional layers, so the same feature extractions are performed, but the new top layers produce predictions for the new task. The added top layers must be retrained from scratch, and are commonly initialized with random weights. During the training process, the rest of the network is frozen, so the weights in those layers maintain their initialized values and do not change during training. VGG19, for example, can be modified to be trained for gender recognition using the design shown in figure 2. Note that in these experiments the dense layer sizes have been decreased to shorten training time and lower graphics card memory consumption. The output layer has been reduced to just two neurons – one for a prediction of male and the other for a prediction of female. The ILSVRC (ImageNet Large Scale Vision Recognition Challenge) weights are frozen inside of VGG19 in the first 16 convolutional layers. Using a model's ImageNet weights is known as *general task* pretraining, and can produce surprisingly good results on a wide variety of datasets [27] [16] [28].

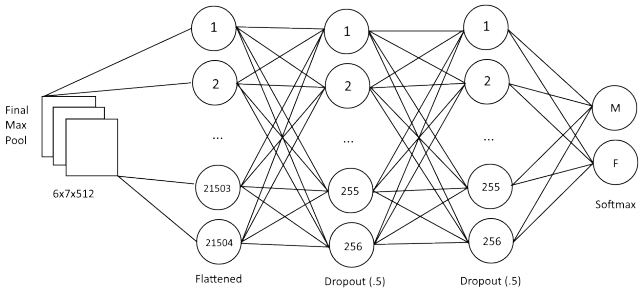

Fig. 2. A new top for VGG19. All weights are initialized randomly and then the network learns how to discern male from female by examining the output of the convolutional layers.

A. Dense Layer Size

There are many factors that can be considered when retraining the top layers of a neural network. An obvious first choice is to test the size and number of fully-connected layers. As seen in figure 3, using only one dense layer seems to inhibit much of the learning process. Loss decreases more quickly when two dense layers are used. This gives traction the argument that more neurons will lead to a higher accuracy. A lower loss does not exactly equate to better validation results, but the general trend is that accuracy increases as loss decreases. The largest top layers, 2048x2 (two dense layers of 2048 neurons) and 4096x2, take several epochs before loss starts decreasing. This effect could be because of the randomly initialized weights. When the top layers are initialized with random weights, it takes more time for large layers to adjust and begin fitting the data. Much smaller dense layers, such as 16x2 and 32x2, also work for gender recognition with about 95% accuracy. Age estimation requires larger dense layers than gender to produce good results.

B. Epochs

Another commonly explored training parameter is the number of epochs for which to train a model. An *epoch* is one pass through all of the data in the training set. Depending on the dataset, depth of the network, regularization techniques, and a variety of other factors, an optimal number of epochs might be high or low. Too few epochs and the network will

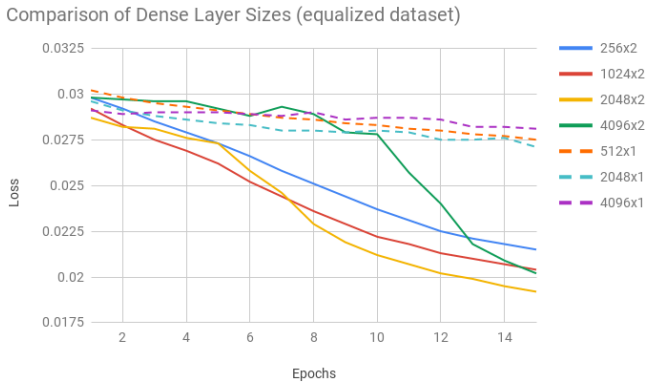

Fig. 3. These are the losses on a short training test using different sizes and numbers of fully-connected layers. The models were trained on 20,000 images and validated on 5,000.

be underlearned. Too many epochs and the model becomes overfit. In both of these instances, validation loss will be higher than usual, and it is unlikely that a near-optimal model will be produced. Figure 4 below shows the $S_2 \cup S_3$ test results at increasing numbers of epochs. For this test, the optimal number of epochs seems to be around 90 where an MAE of 4.753 is achieved. ArgMax and expected value are ways of decoding age from the softmax layer of the neural net and are explained in section VII-D1.

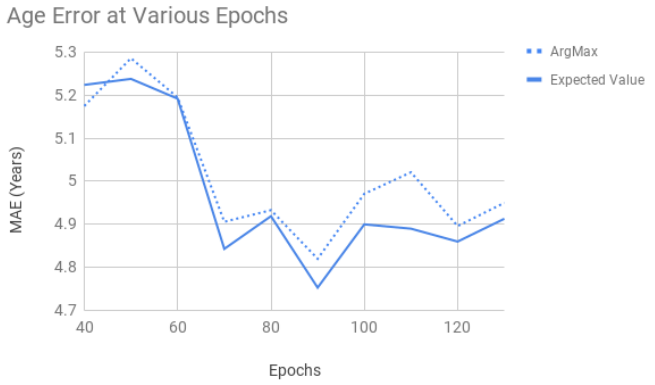

Fig. 4. With epochs, there is a breaking point where a model goes from underlearned to overfit. Usually the best model occurs at that point.

C. Dropout Regularization

One method of combatting overfitting is to add dropout to weight layers. In the 2014 paper that introduces dropout, Srivastava et al. state that it "provides a way of approximately combining exponentially many different neural network architectures efficiently" [29]. When dropout is added to a weight layer, neurons are randomly selected to be removed from the network at each iteration. Those neurons are omitted both when the mini-batch is being fed through the network, and also during backpropagation. The number of neurons removed from the layer is determined by the dropout rate which is set manually. In VGG19, dropout is only used in the top fully-connected layers. Using a higher dropout rate provides more of a regularizing effect, but causes the model to not learn as quickly. In figure 5, dropout can be seen preventing overfitting as the training set losses stay higher, but the test set losses decrease. In table II, the best and final models are also compared. During training, the model that achieves the lowest loss on the validation set is saved and is considered the "best" model. After the last epoch of the last set of trained images, the final model is saved. The best model usually outperforms the final model except if the lowest loss happens to be obtained very early on in the training process. If the model overfits the validation set, a very good loss and accuracy might be recorded, but when fully tested, the performance is mediocre. In the dropout results table, the best result is seen with lower dropout because it acts like a model that is trained for more epochs than 60 which is optimal in this case.

TABLE II
DROPOUT RESULTS

	0.3	**0.4**	**0.5**	**0.6**	**0.7**
Best	**4.768**	5.002	5.192	4.915	5.056
Final	4.953	4.949	6.065	5.698	5.143

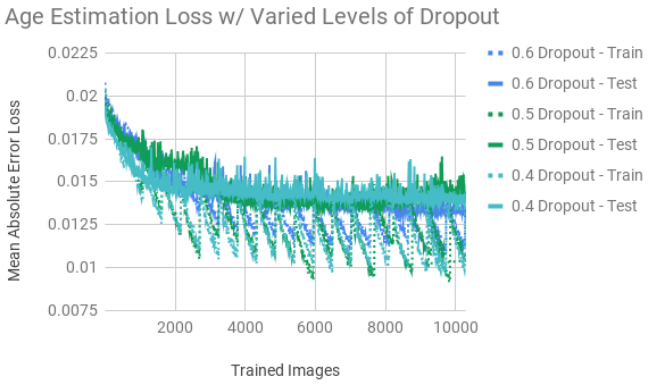

Fig. 5. With lower dropout, the validation loss can be seen to improve more quickly, but it does not reach the depths of the losses that occur at a higher dropout rate. Especially towards the end of the training process, higher dropout can be seen achieving lower losses.

VII. Training Techniques

In addition to changes in the network, many different training techniques were compared in order to observe their effects on gender recognition and age estimation. Training techniques are ways of training a model that can result in better accuracy. Improvements due to the training techniques explored in this section do not result from changes in the network, but from changes in the data.

A. Input Standardization

When VGG was originally submitted to the ImageNet competition, the creators trained on images that had been *zero-centered*. This means that the average was calculated for the training set and subtracted from each pixel value before being fed into the network. Like zero-centering, standardizing image data also centers it at zero, but additionally gives the pixel values a normal spread. When standardizing training data, the validation and test data must also be standardized with respect to the training data. The mean $\bar{x}$, and standard deviation σ, of S_1 can be seen in table III. Once they have been calculated, the formula:

$$\frac{P_i - \bar{x}}{\sigma} \tag{2}$$

can be applied to P which is the set of all pixel values (red, green, and blue) in S_1. As seen in figure 6, standardizing the input data produces immediately better results. Not only does the model begin to fit the dataset faster, it also reaches a higher accuracy than the zero-centered dataset. The accuracies shown in figure 6 are validation set accuracy during each epoch of training. Once trained, the standardized model reaches a gender classification accuracy of 96.209% on the full $S_2 \cup S_3$ test set. This is 1.083% higher than the performance of the model trained on a zero-centered S_1.

TABLE III
S_1 MEAN AND STANDARD DEVIATION

	$\bar{x}$	σ
S_1	142.46	59.85

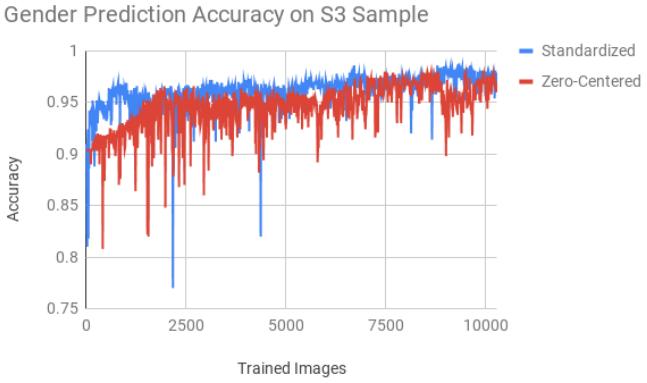

Fig. 6. Standardizing the input helps VGG19 reach higher accuracy faster.

B. Data Augmentation

Data augmentation techniques are commonly used to train neural networks [30] [31]. Since large and accurate datasets are rare and usually private, data augmentation can be used to create more data with which to train a network. For this project, 12-crop resampling was tested. This involves taking a crop from each corner of the image and the center of the image, and resizing the image down to crop size. These six samples are then flipped horizontally to produce twelve unique images. Figure 7 shows an example of a MORPH-II image after 10-crop resampling has been applied (12-crop resampling without the resized image). Using this technique and the training parameters described in section V, an MAE of 5.028 years was obtained. This is a slight increase in performance over the 5.192 MAE of the baseline test, but a drawback is that the network must be trained on 123k 160x200 images instead of 10k 200x240 images, taking 8 times as long to train.

Fig. 7. The original 200x240 images are cropped and flipped to become ten 160x200 images.

C. More Data

Many suggest that a large dataset is integral to deep learning. A sales pitch for deep learning is that deep neural networks can learn more from the data and thereby surpass traditional statistical methods. To test the effects of a larger dataset, the training data are doubled in size by making use of S_2. $S_1 \cup S_2$ becomes a set of 20,560 images and is trained upon, while S_3 is used for testing. The sales pitch appears to ring true as the model achieves an MAE of 4.690 years. Part of the drop in MAE is due to the drop in female population. In $S_2 \cup S_3$, females make up 13.1% of the population but in S_3 the female population is 9.6%.

D. Label Distribution Age Encoding (LDAE)

LDAE is a method of encoding age that has proven more effective than simple one-hot encoding [16]. LDAE recognizes that people age differently, so it helps to view a person's age, denoted by A, as a small scope of potential ages rather than just a binary truth. In this method, the formula:

$$f(i|A,\alpha) = \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{(i-A)^2}{2\alpha^2}} \tag{3}$$

is used to calculate a probability at each age to encode age labels. In the formula A is the age label, i is the age for which a probability should be produced, and α is a hyperparameter that affects the spread of the age probabilities.

1) ArgMax and Expected Value: Two ways that an age can be decoded from the output of a neural network are known as ArgMax and expected value. ArgMax uses the age that has the highest probability. Expected value multiplies the probability at each age by the age and then sums the products. In most

cases, expected value gives more accurate predictions, but the age MAEs are usually fairly close.

2) Dynamic LDAE: In general, it is easier to mistake an old person's age by a large amount than a young person's age. To represent the differing certainty between young and old, the α value in Equation 3 is increased linearly as age increases. In this paper, dynamic LDAE is proposed as follows: an overall α of 2.5 is considered, with a higher α for old ages and a lower α for young ages. To illustrate: During training, age labels are encoded with LDAE from the ages of 5 to 85. The resulting input vector has 81 dimensions, each containing a probability for the corresponding age. For example, considering α with a range from 1 to 4, and for age A, the dynamic α is estimated by:

$$g(\alpha|A) = \frac{range(\alpha)}{81} * (A - 5) + 1. \quad (4)$$

An illustration of the input vector encodings can be seen in figure 8. It improves the accuracy as seen in table IV.

TABLE IV
DYNAMIC LDAE RESULTS

	One-Hot	α = 2.5	α = 1-3.5	α = 1-4
MAE	5.250	5.192	4.861	**4.778**

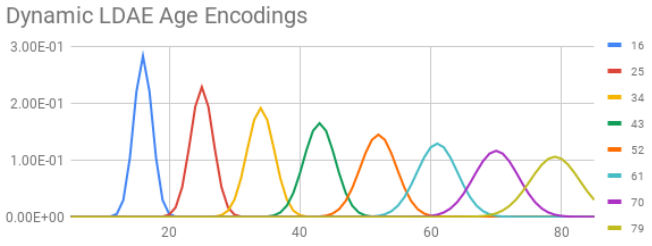

Fig. 8. Training image age labels are encoded with dynamic LDAE. Eight age encodings are shown as α ranges from 1 to 4 and age ranges 16 to 79.

VIII. RESULTS WITH VGGFACE

All of the tests from sections VI and VII use VGG19 with its ILSVRC weights. This section considers transfer learning with the VGG16 architecture pretrained for facial recognition. Although VGG19 is capable of detecting the subtlest differences to separate millions of images into 1000 classes, some of the filters it has learned activate on mundane objects or animal fur, so they do not produce strong activations on images containing human faces. Since VGGFace was originally trained to separate several captures of 2662 individual faces [24], every filter in VGGFace is geared towards finding human facial features. The models in this section were trained using most of the same parameters as above and no extra data augmentation or generalization techniques. As can be seen in tables V and VI, VGGFace takes far fewer epochs to fit the training data even though it is a smaller network. Gender validation set accuracy, for example, reaches 98% during the first epoch of S_1. Because of this, smaller increments of epochs must be searched to find the best resulting models. All models in this section are still trained on S_1 and tested on $S_2 \cup S_3$. As expected, VGGFace produces better results than a general task VGG19 network.

TABLE V
VGGFACE EXPECTED VALUE MAE

Epoch	**5**	**10**	**15**	**20**	**25**	**30**	**35**
MAE	4.800	4.483	4.443	4.468	**4.322**	4.377	**4.323**

TABLE VI
VGGFACE GENDER PREDICTION ACCURACY

Epoch	**3**	**6**	**9**	**12**	**15**	**18**	**21**
Accy.	98.59%	98.47%	**98.68%**	98.53%	98.65%	98.64%	98.56%

IX. GENDER-SPECIFIED HIERARCHAL AGE MODEL

In the past, hierarchies of classifiers have used multiple feature labels of datasets to increase classification accuracy. Test data are separated into different classes before being classified again by models trained specifically for each class. Guo and Mu use a hierarchy of KPLS (kernel partial least squares) race and gender classifiers with BIFs to obtain a MORPH-II MAE of 4.18 years [22]. It makes sense that men and women would have different features that are indicative of age, so a hierarchy of deep CNN models might also produce better age estimation results than a single model used for both genders. Because female image data are fairly limited in MORPH-II, an 80/20 train/test split was devised such that no subjects who appear in the training data are also in the testing data. A gender model was trained for the hierarchy using the same 80% portion of the training data and it achieved an accuracy of 98.60%.

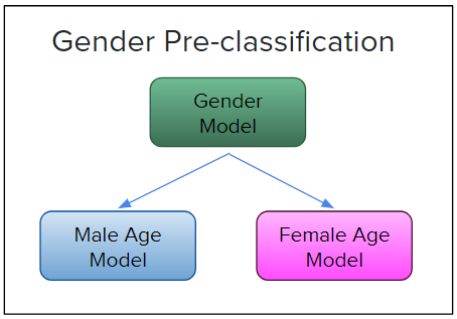

Fig. 9. In this design, images are classified as male or female before estimating age with the corresponding model.

The female age model achieved an MAE of 5.22 years after being trained for 25 epochs on all female images in the training set. The male age model achieved an MAE of 3.79 years after 20 epochs. The deep CNN hierarchy achieved an

MAE of 4.10 years which outperforms all other age estimation models in this document. This experiment shows that with enough training data, a hierarchy of deep CNNs can surpass a single model trained for general age estimation. Additionally, the model can make age and gender predictions at a rate of 62.62 frames per second meaning that it is suitable for real-time age and gender estimation deployment. Table VII shows a comparison of the results from this paper alongside other MORPH-II studies.

TABLE VII
MORPH-II RESULTS WITH COMPARISON

Approach	Year	Train	Test	Age	Gender
BIF+OLPP [32]	2010	S_1, S_2	$S_2 \cup S_3$, $S_1 \cup S_3$	4.45	97.84%
BIF+KPLS [22]	2011	S_1, S_2	$S_2 \cup S_3$, $S_1 \cup S_3$	4.18	98.20%
VGG19	2018	S_1	$S_2 \cup S_3$	4.75	96.6%
VGGFace	2018	S_1	$S_2 \cup S_3$	4.32	98.68%
Gender-Specified	2018	80%	20%	4.10	98.60%

X. CONCLUSION

Although VGG19 was not originally trained to recognize faces, good results for gender recognition and age estimation can still be obtained using transfer learning techniques. Transfer learning with a pretrained model that is more pertinent to the task, such as VGGFace, can produce results that beat other gender recognition and age estimation techniques, and can even exceed human performance. Changes in network designs and training techniques can be studied without having to spend weeks training models from scratch. The models for this paper were all trained using a GTX 1060 Max-Q and a GTX 1070. This paper has shown the advantages offered by certain model designs, training techniques, and pretrained weights. It has also demonstrated that hierarchies of AI models offer promise and should be considered when implementing a classification system.

Future Work The results shown here surpass nearly all results obtained before 2012 simply by leveraging new deep learning technology. They are, however, far from what is possible. A live demo of these AI models revealed flaws. Age estimations were occasionally wildly off for minorities in the dataset such as Asians, Hispanics, women, children, and elderly people. Additionally, gender predictions seemed largely based on the absence or presence of long hair and could change by the tilt of the head. To address these types of issues further training techniques could be studied that help to generalize and stabilize prediction accuracy. Larger and more balanced datasets could be used as training data. Newer types of deep CNN architectures could be adapted to age and gender estimation and might yield better results even with general task weights. Increasingly deep hierarchies of models can be considered for appropriately large datasets. Combinations of datasets and fusions of features could be what's on the horizon for the advancement of deep learning. Finally, training models from scratch specifically for age and gender predictions would probably produce better results.

XI. ACKNOWLEDGEMENTS

This research was conducted with funding from the National Science Foundation under DMS Grant Number 1659288.

REFERENCES

[1] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai *et al.*, "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018. 1

[2] A. Prieto, B. Prieto, E. M. Ortigosa, E. Ros, F. Pelayo, J. Ortega, and I. Rojas, "Neural networks: An overview of early research, current frameworks and new challenges," *Neurocomputing*, vol. 214, pp. 242–268, 2016. 1

[3] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, 2017. 1

[4] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016. 1

[5] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015. 1

[6] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in *Proceedings of the 16th ACM International Conference on Multimedia*, ser. MM '08. New York, NY, USA: ACM, 2008, pp. 725–728. [Online]. Available: http://doi.acm.org/10.1145/1459359.1459470 1

[7] G. Guo and G. Mu, "A framework for joint estimation of age, gender and ethnicity on a large database," *Image and Vision Computing*, vol. 32, no. 10, pp. 761 – 770, 2014, best of Automatic Face and Gesture Recognition 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885614000791 1

[8] T. Wilhelm, H.-J. Böhme, and H.-M. Gross, "Classification of face images for gender, age, facial expression, and identity," in *Artificial Neural Networks: Biological Inspirations – ICANN 2005*, W. Duch, J. Kacprzyk, E. Oja, and S. Zadrożny, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 569–574. 1

[9] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," *2013 International Conference on Biometrics (ICB)*, 2013. 1

[10] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, and et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, p. 211–252, Nov 2015. 1

[11] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010. 1

[12] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*. IEEE, 2006, pp. 341–345. 1, 2

[13] R. Rothe, R. Timofte, and L. V. Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision (IJCV)*, July 2016. 1

[14] Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, and S. Yan, "Facial age estimation with age difference," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3087–3097, 2017. 1

[15] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output cnn for age estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4920–4928. 1

[16] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective Training of Convolutional Neural Networks for Face-based Gender and Age Prediction," *Pattern Recognition*, vol. 72, p. 15–26, 2017. 1, 2, 3, 5

[17] M. Castrillón-Santana, J. Lorenzo-Navarro, and E. Ramón-Balmaseda, "Descriptors and regions of interest fusion for in-and cross-database gender classification in the wild," *Image and Vision Computing*, vol. 57, pp. 15–24, 2017. 1

[18] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 34–42. 1

[19] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal Regression With Multiple Output CNN for Age Estimation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016. 1

[20] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15–26, 2017. 1, 2

[21] X. Geng, C. Yin, and Z.-H. Zhou, "Facial age estimation by learning from label distributions," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 10, pp. 2401–2412, 2013. 1

[22] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," in *CVPR 2011*, June 2011, pp. 657–664. 1, 2, 6, 7

[23] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556 2

[24] O. M. Parkhi, A. Vedaldi, and A. Zisserman, ""Deep Face Recognition"," in *British Machine Vision Conference*, 2015. 2, 6

[25] G. Bingham, B. Yip, M. Ferguson, and C. Nansalo, "MORPH-II: Inconsistencies and Cleaning Whitepaper," Oct 2017. [Online]. Available: http://libres.uncg.edu/ir/uncw/f/wangy2017-1.pdf 2

[26] B. Yip, R. Towner, T. Kling, C. Chen, and Y. Wang, "Image Pre-processing Using OpenCV Library on MORPH-II Face Database." [Online]. Available: https://uncw.edu/math/REU/documents/image-pre-processing.pdf 3

[27] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3320–3328. [Online]. Available: http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks.pdf 3

[28] S. M. Xie, N. Jean, M. Burke, D. B. Lobell, and S. Ermon, "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping," *CoRR*, vol. abs/1510.00098, 2015. [Online]. Available: http://arxiv.org/abs/1510.00098 3

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html 4

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf 5

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going Deeper with Convolutions," *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: http://arxiv.org/abs/1409.4842 5

[32] G. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?" in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, 2010, pp. 71–78. 7