HARNESSING UNLABELED DATA FOR IMPROVING GENERALIZATION OF DEEP LEARNING METHODS

A Thesis by

Deepika Shanmugasundaram

Bachelor of Engineering, Anna University, 2020

Submitted to the School of Computing
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

July 2023

HARNESSING UNLABELED DATA FOR IMPROVING GENERALIZATION OF DEEP LEARNING METHODS

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Computer Science.

Ajita Rattani, Committee Chair

Rajiv Bagai, Committee Member

Sergio Salinas Monroy, Committee Member

Kaushik Sinha, Committee Member

Zelalem Demissie, Committee Member

ACKNOWLEDGEMENTS

I would like to express my gratitude and deepest appreciation to my advisor, Dr Ajita Rattani, for her guidance throughout my program. I extend my gratitude to members of my committee, Dr Rajiv Bagai, Dr Sergio Salinas Monroy, Dr Kaushik Sinha and Dr Zelalem Demissie for their suggestions and comments on this research work.

ABSTRACT

Recent advancements in Deep Learning, Artificial Intelligence, and Computer Vision have reached a critical stage, enabling researchers to explore the automatic extraction of individual demographic traits, known as soft-biometrics. This research aims to leverage unlabeled data in predicting soft-biometric traits, such as gender and age, using deep learning models. The objective is to develop a model that can accurately classify these traits by utilizing semi-supervised methods that rely on a limited amount of labeled data and a vast amount of unlabeled data. While unlabeled data may initially seem devoid of crucial information, this thesis explores how it can be effectively used to enhance classification accuracy, especially in scenarios where labeled data is scarce.

This study evaluated the accuracy of different image classification models on the Celeb-A and NIR-VIS datasets using co-training, mix-up procedure, knowledge distillation, and blind distillation techniques. The results showed that incorporating these methods led to improvements in accuracy across both datasets and various attributes such as gender classification and smiling classification. Exploring the combined use of different techniques and investigating their synergistic effects could lead to further accuracy improvements. Evaluating the models on larger and more diverse datasets, analyzing their generalization capabilities, optimizing hyperparameters and architectures, and applying the techniques to other computer vision tasks were also identified as areas for future research.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF TABLES

LIST OF ABBREVIATIONS

NIR : Near-Infrared Spectrum

VIS : Visible Spectrum

CNNs : Convolution Neural Networks

BANNs : Born-Again Neural Networks

KL-Div : Kullback–Leibler Divergence

CHAPTER 1

1 INTRODUCTION

1.1 Problem Statement

The improvements in Deep Learning, Artificial Intelligence and Computer vision have reached a critical point. A variety of research have been conducted to investigate the automatic extraction of an individual's demographic traits such as gender, age, ethnicity, and so on, which is known as soft-biometrics[1][2]. In traditional approaches to identifying soft- biometrics, much reliance is placed on a huge scale of labeled data, which is generally difficult and expensive to gather. Deep learning methods necessitate a significant amount of training data in order to achieve high generalization accuracy rates. However, acquiring huge biometric datasets as well as soft-biometric variables such as gender, race, and age may raise privacy and security concerns.

As a result, Semi-supervised learning is one solution to this challenge [3]. Semi-supervised learning requires combining labeled and unlabeled data to improve the classifier's generalization performance. Semi-supervised learning algorithm such as co-training have received a lot of interest. Co-training, first proposed by Blum and Mitchell [4], is another prominent semisupervised learning paradigm. It is typically applied to datasets that have a natural separation of their features into two distinct groups, which are viewed as two

views of the data.

1.2 Background And Motivation

High image classification accuracy rates are acquired at the expense of a large number of parameters, which necessitates large scale labeled datasets for network training and finding the ideal set of parameters. The availability of large-scale labeled datasets may be limited due to the need for manual annotation by a human expert, which is a tedious and costly activity.

Unlabeled data is substantially simpler to get than in traditional semi-supervised or transfer learning methods. High-accuracy image classification can be achieved using semi-supervised learning algorithms that use a limited number of labeled samples and a large number of inexpensive unlabeled samples.

1.3 Research Objective

The objective of this research is to make use of unlabeled data in the predictions of soft-biometrics[1][2] in different biometrics traits using deep learning models. The research tries to create the model which can classify the gender and age using semi-supervised methods which use limited label data and huge unlabeled data. At first look, it may appear that unlabeled data provides no benefit. After all, an unlabeled data lacks the most critical piece of information in its class. In this thesis, We explain how unlabeled data can be used and utilized to improve classification accuracy, particularly when labeled data is scarce. Different deep learning models, such as co-training[4]

with Mix-up procedure [5], Blind Distillation[6], Augmenting both labels[7] etc. are being developed. Co-training [4] technique is used to classify the features of images into two different views. In co-training, two-classifiers work together to generate pseudo-labels and re-train classifiers on unlabeled datasets.

1.4 Gender Classification Using Deep Learning Methods

It can be advantageous to infer features such as the gender or age of the people involved when examining human behaviors using data mining or machine learning approaches. Gender[8][9][10] and age[8][9][10]have all been deduced from facial images in studies on face biometrics. we have several models to classify gender and age groups.

1.4.1 Image Processing

Image processing computerization is commonly used for face recognition, expression recognition, age determination, racial binding, and gender categorization. Gender classification is simple for us because we can tell whether a person is male or female based on their hair, nose, eyes, mouth, and skin with a high degree of confidence and accuracy; however, can we program a computer to perform just as well at gender classification? Face identification, noise removal, face alignment, feature representation, and classification are the five steps in the traditional sequence for contemporary real-time facial image processing.

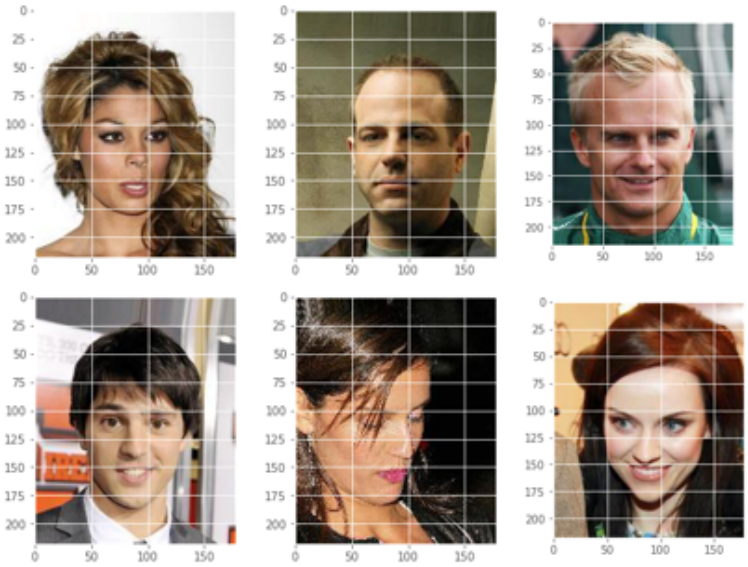

Figure 1: Sample images from Celeb-A Dataset

1.4.2 Celeb-A Dataset

CelebFaces Attributes Dataset (CelebA)[11] is freely available dataset. It is a large-scale face attributes dataset with over 200K celebrity photos with 40 attribute annotations. The images in this collection span a wide range of position variants as well as background clutter. CelebA offers a wide range of annotations, as well as a big number of them.

1.4.3 NIR-VIS Dataset

The most typical case of heterogeneous face recognition is near-infrared to visible (NIR-VIS)[12] face recognition, which seeks to match a pair of face images taken from two different modalities.

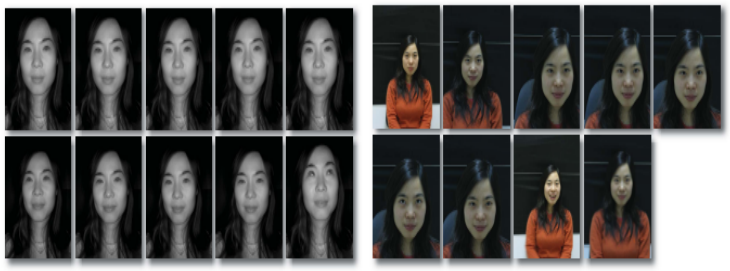

Figure 2: Sample images from NIR-VIS Dataset

1.4.4 Co-Training

Co-training[4] trains two learners from two separate views and allows the learners to label the most confident unlabeled occurrences in order to expand the training set of the other learner. This method can be performed indefinitely until a halting condition is reached.

1.4.5 Knowledge Distillation

Knowledge distillation[13] is a technique used in machine learning to transfer knowledge from a large, complex model (known as the teacher model) to a smaller, more efficient model (known as the student model). The goal of knowledge distillation is to distill the knowledge learned by the teacher model into the student model, enabling it to achieve similar performance with less computational resources. The process of knowledge distillation involves training the student model to mimic the behavior of the teacher model. This is typically done by using the outputs of the teacher model, often in the form of soft targets or probability distributions, as additional supervision during the training of the student model. By incorporating the knowledge

from the teacher model, the student model can learn to generalize better and make more accurate predictions. The teacher model is usually a larger and more powerful model that has been trained on a large dataset. It can be a state-of-the-art model, such as a deep neural network or a complex ensemble of models. The student model, on the other hand, is designed to be more lightweight and computationally efficient, making it suitable for deployment in resource-constrained environments, such as mobile devices or embedded systems. Overall, knowledge distillation is a powerful technique that allows for the transfer of knowledge from large, complex models to smaller, more efficient models, enabling them to achieve similar performance with reduced computational resources.

1.5 Contributions Of Thesis

The main contributions of this study is developing deep learning models to classify the soft-biometrics[1][2] using the unlabeled samples. Mainly focusing on Utilizing Unlabeled data to improve the quality of the model which is considered to be not applicable to classify images.

CHAPTER 2

2 PRIOR STATE OF THE ART TECHNIQUES

2.1 Co-training

To make use of unlabeled data, many state of the art techniques were employed. Semi-supervised learning, which employs both labeled and unlabeled data to train a classifier, has an extensive background in machine learning research[3]. The seminal work by Blum and Mitchell[4] used the naive Bayes classifier as a base learner. Subsequent research has explored various classifier combinations, such as decision trees, support vector machines, and neural networks. The algorithms are discussed in terms of their training process, instance selection strategies, and convergence criteria. Several research studies have focused on extending and refining the co-training framework, as well as exploring its limitations and potential improvements. Nigam et al. (2000)[14] extended co-training by introducing a framework for handling data with multiple types of attributes. They applied their approach to text classification tasks and showed significant improvements over single-view learning. Since the publication of Blum and Mitchell's paper[4], their co-training approach has been extensively studied and applied in various domains. Zhang and Zhou (2007)[15] proposed a self-training co-training algorithm that re-

duces the dependency on initial labeled data by iteratively expanding the labeled set through the agreement of the multiple classifiers. They demonstrated its effectiveness on both text and image classification tasks. Cheng et al.(2018)[16] introduced a deep co-training framework for semi-supervised image recognition. They combined deep learning models with co-training, utilizing unlabeled data to improve the performance of deep neural networks. The proposed approach achieved competitive results on several benchmark datasets, demonstrating the effectiveness of co-training in the context of deep learning-based image classification.

2.2 Mix-up Procedure

H.Zhang et al[5] introduce the mixup regularization technique, which goes beyond empirical risk minimization in training deep neural networks. mixup involves creating virtual training samples by linearly interpolating between pairs of labeled examples and their corresponding labels. The authors demonstrate that mixup improves generalization and robustness of neural networks across various tasks and datasets. Tokozume and Harada[17] propose a variant of mixup called between-class learning, which further extends the concept of mixup. This approach creates virtual training samples not only between pairs of labeled examples but also between different classes. The authors show that between-class learning enhances the discriminative capability of deep neural networks and achieves state-of-the-art results on image classification tasks.

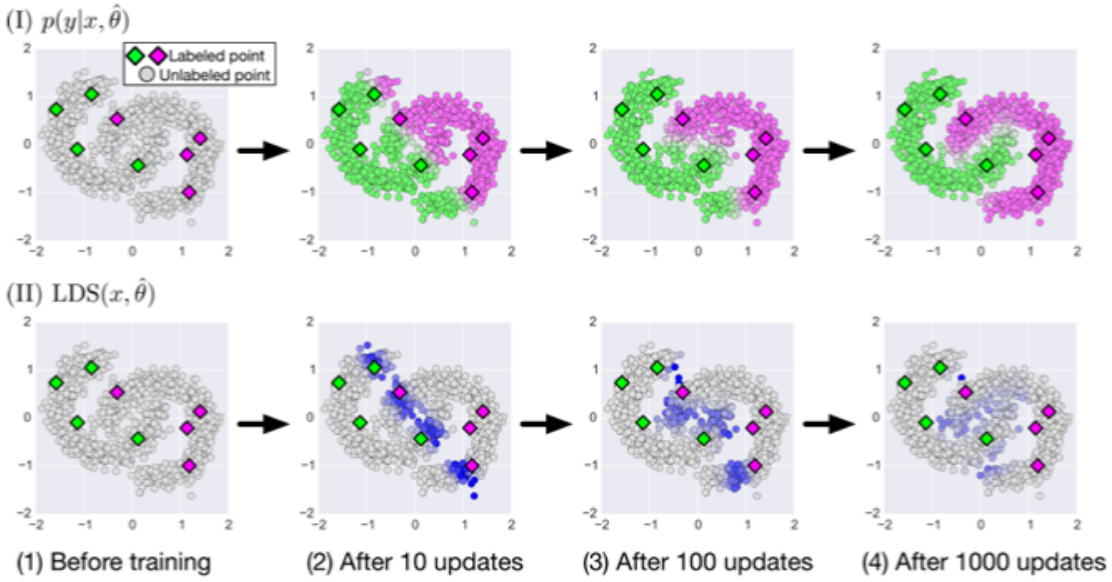

Figure 3: Demonstration of how our VAT works on semi-supervised learning. It created 8 labeled data points in 2-D space (y=1 and y=0 are green and purple, respectively) and 1,000 unlabeled data points. The panels in row I show the prediction p(y=1—x,) on the unlabeled input points at various phases of the algorithm. It employed a continuous colormap to represent the expected values of p(y=1—x,), with green, gray, and purple corresponding to the values 1.0, 0.5, and 0.0, respectively. The heatmaps of the regularization term LDS(x,) on the input points are shown in the second row (II). LDS values on blue-colored points are relatively high as compared to gray-colored ones. It made advantage of KL divergence.

2.3 Augment Both Label and Unlabeled

Miyato et al[7] introduces virtual adversarial training (VAT) as a regularization technique for improving the robustness and generalization of deep neural networks. The authors propose a method to generate small adversarial perturbations that maximize the Kullback-Leibler (KL) divergence between the predictions on original and perturbed inputs. VAT is demonstrated to achieve state-of-the-art performance in supervised and semi-supervised learning scenarios. Miyato et al[18] extend VAT to the domain of text classification. They apply adversarial training techniques to train text classification models using both labeled and unlabeled data. The proposed method achieves improved performance compared to other semi-supervised learning approaches. Saito et al[19] build upon the VAT framework and propose asymmetric tri-training, a method for unsupervised domain adaptation. The authors use three classifiers and iteratively update their parameters using unlabeled target domain data. VAT is employed as a regularization method during the training process, leading to improved domain adaptation performance.

2.4 Knowledge Distillation

Hinton et al[13] introduced the concept of knowledge distillation and proposed a method to transfer the knowledge from a large teacher network to a smaller student network. It discusses the use of soft targets and explores the benefits of knowledge distillation. Romero et at[20] where the student model is trained to mimic the intermediate representations of the teacher

model. This paper presents experimental results showing the effectiveness of FitNets in knowledge distillation. The work extends knowledge distillation to focus on attention mechanisms in convolutional neural networks (CNNs). The authors propose a method to transfer the attention maps of the teacher network to the student network, leading to improved performance[21]. Chen G et at[22] explores the application of knowledge distillation in the context of object detection. The authors propose a method to transfer the knowledge from a large object detection model to a smaller model, achieving comparable performance with reduced computational requirements. Heo J et al[23] introduces activation-based knowledge distillation, where the teacher and student models are trained to match the activation boundaries formed by hidden neurons. Experimental results demonstrate the effectiveness of this approach in various tasks. The Born-Again Neural Networks (BANNs), a method that combines self-distillation and knowledge distillation. The student model is initially trained on the training data, and then the student model is used as a teacher to distill knowledge back into itself, resulting in further improvements in performance[24].

2.5 Blind Distillation

Gaurav et al[6] proposes a term Blind Distillation, It is used in the context of machine learning and knowledge transfer, specifically in the field of neural network models. It refers to a technique where a smaller or "student" neural network is trained to mimic the behavior and predictions of a larger or

"teacher" neural network without having access to the labeled training data that was originally used to train the teacher network. They proposed training a smaller "student" network to mimic the behavior of a larger "teacher" network by using the soft target probabilities generated by the teacher network. The approach demonstrated improved performance and generalization of the student network[13]. Zagoruyko and Komodakis[21] proposed attention transfer, a technique based on blind distillation, to improve the performance of convolutional neural networks (CNNs). They transferred the attention maps of the teacher network to guide the learning of the student network, resulting in improved accuracy and robustness.

CHAPTER 3

3 METHODOLOGY

3.1 Proposed methods for classification

Proposed method talks about classification of gender and smile detection using unlabeled samples. The methods to reduce the usage of labeled images in the models so that they can be less expensive for classification. The models proposed for classification are co-training incorporated with mix-up procedure, augmenting both label and unlabeled, Knowledge distillation and blind distillation. To classify images co-training[4] method is used. Celeb-a and NIR-VIS datasets are used to classify images. The CelebA[11] dataset is a large-scale face attributes dataset that is commonly used in computer vision and machine learning research. It consists of over 200,000 celebrity images with annotations for 40 different attribute labels per image. . Notre Dame Near-Infrared and Visible-Light (NDNIVL) [12]is the largest available NIR/VIS dataset with 24,605 images of 574 subjects. To use unlabeled images co-training methods were used wisely to obtain high classification accuracy. On the basis of their accuracy, the models are compared against current state-of-the-art models.

Table 1: Data split of Datasets into train, test and validation

Data Split	Celeb-A Image Pairs	NIR-VIS Image Pairs
Train	162770	18039
Test	19867	3744
Validation	19963	2338

3.2 Celeb-A Dataset

he CelebA dataset[11] is a large-scale face attributes dataset that is commonly used in computer vision and machine learning research. It consists of over 200,000 celebrity images with annotations for 40 different attribute labels per image. These attributes include binary labels such as "Male" or "Female" as well as more specific attributes like "Eyeglasses," "Smiling," and "Young." The dataset was created by researchers at the Chinese University of Hong Kong (CUHK) and is widely used for tasks like facial recognition, attribute prediction, and face generation. It provides a valuable resource for training and evaluating algorithms in the field of computer vision and machine learning. Each image in the CelebA dataset is of size 178x218 pixels and contains a single face. The dataset is known for its diversity in terms of age, gender, and ethnicity, making it suitable for various applications and research purposes.

The images in the dataset has been classified into train, test and validate. They are represent as 0-train, 1-test, 2-validate. There are 162770 training images, 19867 test images and 19963 validate images.

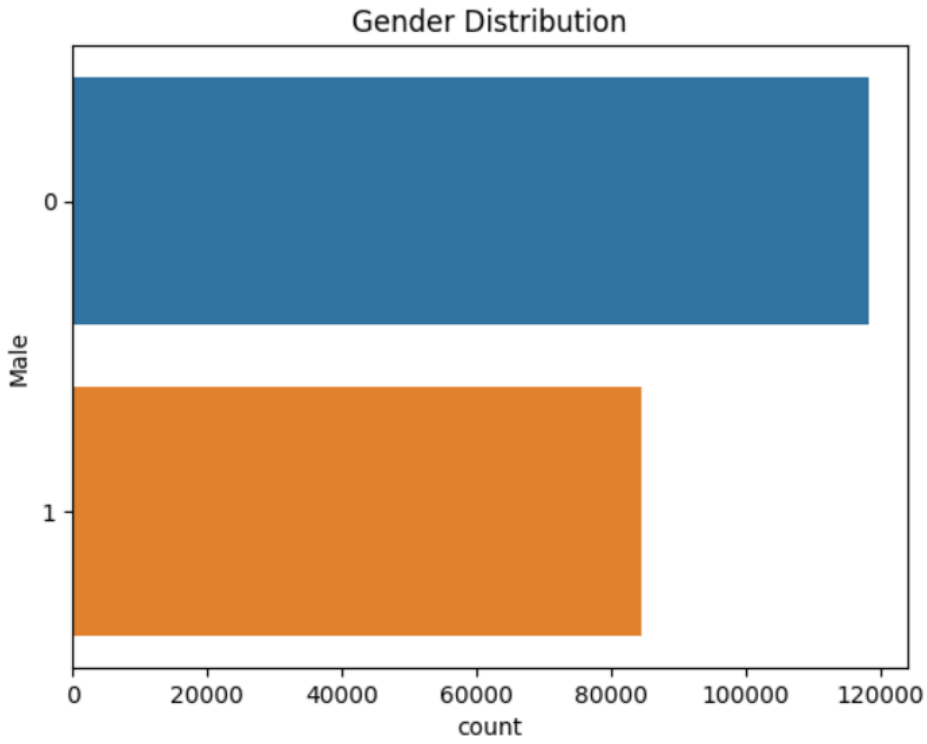

Figure 4: Gender Distribution in Celeb-A Dataset

3.3 NIR-VIS Dataset

The Notre Dame Near-Infrared and Visible-Light (NDNIVL)[12] collection contains 24,605 images of 574 people, making it the biggest known NIR/VIS dataset. All images were taken in natural indoor lighting with a frontal position and neutral facial expression. The resolution of the NIR pictures is 4770x3177. As a result, ND-NIVL now has the biggest archive of high-resolution NIR photos. The dataset has classified into train, test and validate. There are 18039 train images, 3744 test images and 2338 validate images.

3.4 Implementation Details

All the experiments were performed on Intel Xeon(R) Platinum 8268 CPU @ 2.90GHz x96. Each model was trained for the ratio of total train images to batch size . The learning rate is 0.00004 and the models were trained on batch size of 32. The loss function used for all models was binary cross entropy. Adam is the optimizer that we used for all models. For all models images of size 224 x 224 are used.

3.5 Image Classification

A few of models were examined for image classification. The models that are employed are Co-training, Knowledge Distillation, Mix-up procedure, Augmenting both label and unlabeled and Blind distillation. For training all the models used celeb-a dataset and NIR-VIS dataset has been used. To train 162770 images from celeb-a dataset and 18039 images from NIR-VIS

dataset were used. To train 19867 images from celeb-a and 3744 from NIR-VIS dataset were used.

3.5.1 Unlabeled samples on pre-trained models

When training a pre-trained models using unlabeled samples, the training typically occurs in batches. Here's how unlabeled samples can be used in training on a batch-wise manner. Gathering a large set of unlabeled images in dataset that want to use for training. Next, Preprocess the unlabeled images to ensure they are in the appropriate format and size for pre-trained models. This usually involves resizing the images to the required input dimensions and applying any necessary normalization or data augmentation techniques. Divide the unlabeled dataset into batches. The batch size is a hyperparameter that determines the number of unlabeled samples used in each training iteration. The batch size can vary based on computational resources and the size of the unlabeled dataset. For each batch of unlabeled images, perform a forward pass through the pre-trained ResNet50 model. This means passing the images through the model to obtain the model's predictions or feature representations. Extract the relevant features from the pre-trained model's output. Use the extracted features as inputs to update the model's parameters. Repeat the process for the next batch of unlabeled images. Continue iterating through the batches until you have processed all the unlabeled data or have reached a desired number of training steps or epochs. By training on unlabeled samples in batches, you can utilize the large amount of available unlabeled data efficiently. The model learns from

the unlabeled samples to improve its representations and generalization capabilities. These representations can then be further refined using labeled data or used for downstream tasks such as classification or clustering.

3.5.2 Basic Co-training Model

To classify the gender using unlabeled data a basic co-training model is developed. The foremost step is to extract the feature from the data that can be used to differentiate gender. The ResNet50 pre-trained model is used to extract the features from the face images. The ResNet50 models are pre-trained on ImageNet. The features are extracted from ResNet50 model on the average pool layer. Trained the labeled dataset and divided the labeled dataset into two disjoint subsets, each representing different views. One subset will contain the raw face images, the other view will contain the extracted features using the ResNet50 model. Train separate classifiers on each view of the data using the labeled examples. For the view containing raw face images, you can fine-tune the ResNet50 model by adding a few additional layers of Dense(1024, activation = 'relu'), Dense(1, activation = 'sigmoid') and Dropout(0.5) to avoid over fitting. Classifiers are used to train on one view to assign the pseudo labels to the unlabeled samples on other view. Used the fine-tuned ResNet50 model to predict the gender labels for the unlabeled images in the feature view. Assign the predicted gender labels as pseudo-labels to those unlabeled examples. Identified highly classified samples by considering the agreement between the classifiers. Compare the predicted gender labels and their confidence scores across the classifiers. The threshold value

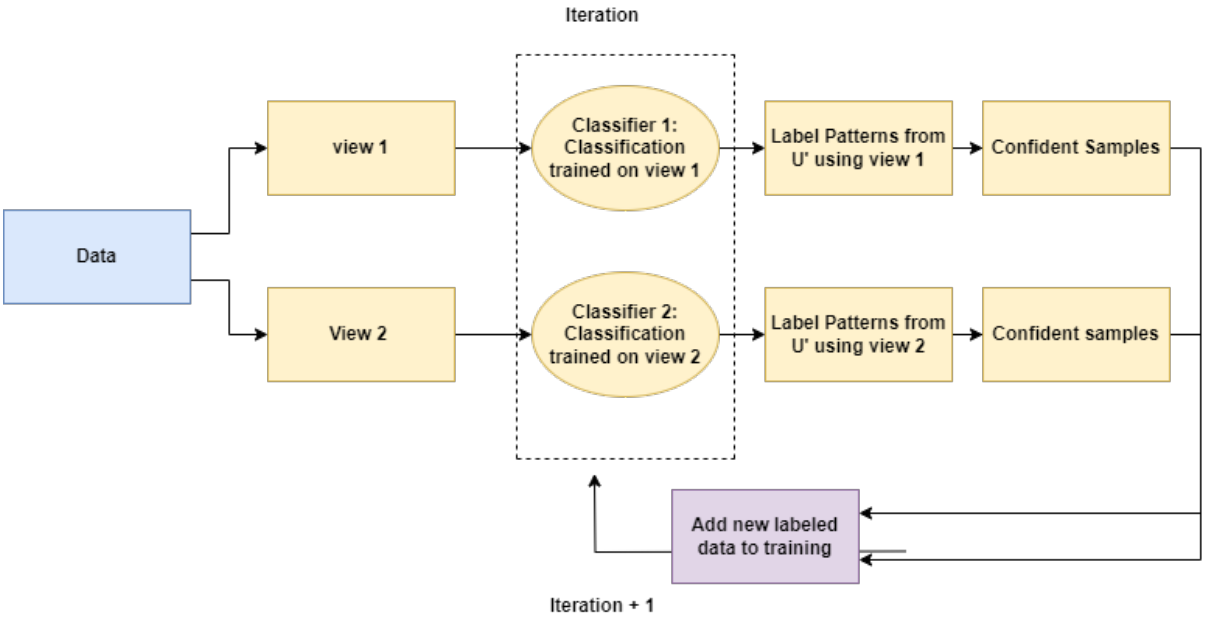

Figure 5: Basic Co-training model

between 0.1 to 0.9 are selected as the samples where the classifiers agree with high confidence. Combined the labeled examples with their pseudo-labels to create an augmented labeled dataset. Include the highly classified samples in this augmented dataset. Repeated the procedure for multiple Iterations, alternating between the views and updating the classifiers and labeled dataset at each iteration. After several iterations, combine the predictions from the classifiers trained on both views using weighted average to obtain the final gender classication. Evaluated the performance of the co-training approach on a separate validation set using accuracy metric.

3.5.3 Co-training with Mix-up Procedure

Combined co-training[4] and the mix-up[5] technique can be an effective approach for training a model. Co-training is a semi-supervised learning technique where multiple models are trained simultaneously on different views of the data. In combination with mix-up, it can further enhance the model's generalization ability and performance. For a given pair of samples (x_1, y_1) and (x_2, y_2), where x_1 and x_2 are the input samples and y_1 and y_2 are the corresponding labels, mixup generates a mixed sample x_{mix} and a mixed label y_{mix} using the following formula:

$$x_{mix} = \lambda x_1 + (1 - \lambda)x_2$$

$$y_{mix} = \lambda y_1 + (1 - \lambda)y_2$$

Here, λ is a randomly generated mixing coefficient from a Beta distribution with parameters α and α. The value of α controls the strength of the mixup. The mixing coefficient ensures that the mixed sample and label are a convex combination of the original samples and labels.

By generating mixed samples and labels using the mixup formula, the mixup technique encourages the model to learn more robust and generalizable representations by exposing it to interpolated samples and labels during training.

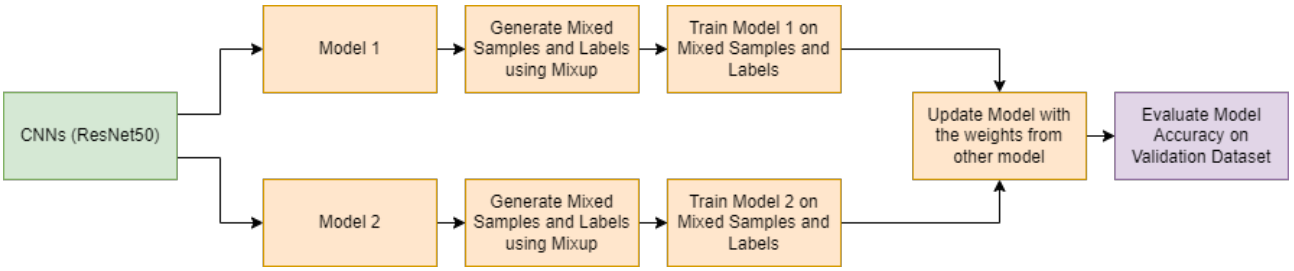

Figure 6: Co-training with Mix-up Procedure

It's implemented using the ResNet50 pre-trained model. It has two models pre-trained with ResNet50 by excluding the top classification layer. Freeze the pre-trained layers for both models. Add a global average pooling layer and a dense layer. Define mixup-loss , the loss will determined by categorical crossentropy. Compiled both models using adam optimizer and mixup-loss. Generated mixed samples and labels using mix-up.The Hyper parameter alpha is 0.2 for mix-up. Then train the both model with mixed images and mixed labels. Updated model 1 with the weights from model 2 and vice versa. Finally evaluated model accuracy on validation dataset.

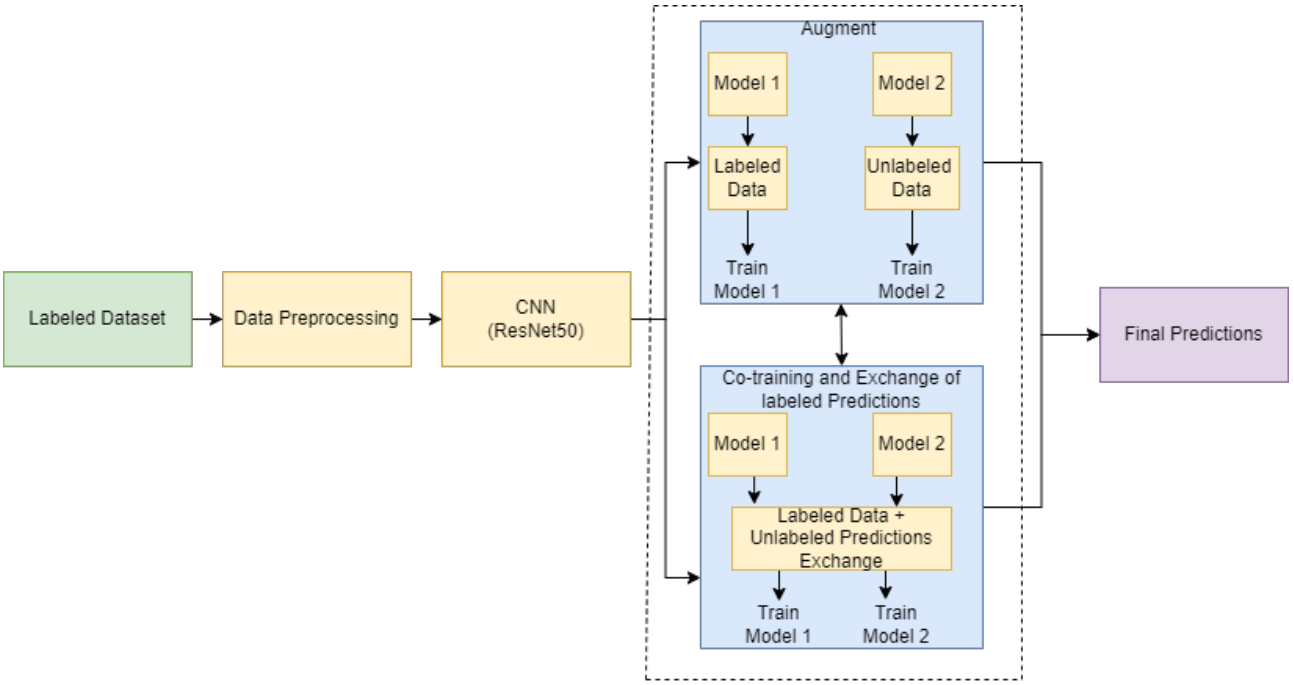

Figure 7: Co-training with Augmenting both label and unlabeled

3.5.4 Co-training with Augmenting both label and unlabeled

Define the custom dataset class for labeled and unlabeled data. Set up transformations for data processing by resized transform 224*224 size and normalized mean is 0.485, 0.456 0.406 and standard deviation 0.229, 0.224, 0.225. Then created dataloaders for labeled and unlabeled data. Defined two ResNet50 model. Defined virtual adversarial loss by using softmax logits and Kullback–Leibler divergence (KL-Div) loss is used. Then trained VAT using virtual adverserial loss and cross entropy loss.Then train Co-training with models, label and unlabeled dataloaders and optimizer adam. The both model uses cross entropy loss for training. Finaly evaluate on validation set.

3.5.5 Knowledge Distillation using Co-training

It initializes the teacher model (ResNet-50) with pre-trained weights and the student model (ResNet-8) without pre-trained weights. Pre-trained weights are used for the teacher model to benefit from the knowledge learned from a Celeb-a and NIR-VIS dataset. The student model starts with random weights and will be trained to mimic the teacher model's behavior. It trains the student model using knowledge distillation. Knowledge distillation is a technique where the soft labels (i.e., the probabilities or logits) from the teacher model are used as targets for training the student model. The student model is trained to produce similar outputs to the teacher model, effectively transferring knowledge from the teacher to the student. The distillation loss is a combination of the Kullback-Leibler (KL) divergence between the student

and teacher outputs and the standard cross-entropy loss. After training the student model using knowledge distillation, the code initializes a co-training model (also ResNet-8) without pre-trained weights. The co-training model will be trained simultaneously with the student model using the same distillation loss. The purpose of co-training is to further enhance the performance of the student model by combining the knowledge learned by two different models. It trains both the student model and the co-training model simultaneously. During each training iteration, the models receive the same input images, and the distillation loss is calculated using the teacher model's soft labels. The loss is a combination of the student loss and the co-training loss, weighted by the beta parameter. This ensures that both models learn from the teacher's knowledge while also benefiting from each other's learned representations. fter the training is completed, the code evaluates the performance of the ensemble model, which consists of the student model and the co-training model. The models are set to evaluation mode (eval()), and the validation dataset is used to calculate the validation loss and accuracy. The ensemble model's outputs are obtained by averaging the outputs of the student and co-training models. This evaluation step provides an assessment of how well the combined models generalize to unseen data. By combining knowledge distillation and co-training, the code aims to improve the performance of the student model by leveraging the knowledge from the teacher model and the co-training model. This allows the student model to achieve better accuracy compared to training with knowledge distillation alone.

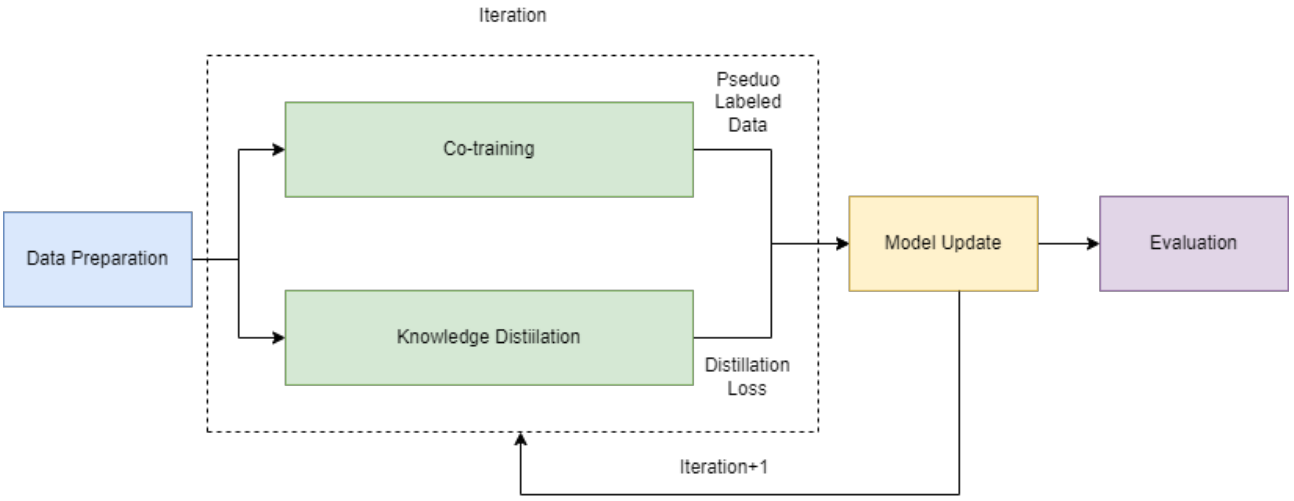

Figure 8: Knowledge Distillation using Co-training

3.5.6 Blind Distillation using Co-training

This model uses a blind distillation[6] and co-training[4] to develop as better model for using unlabeled samples. Transformations are defined using transforms.Compose to resize images to (224, 224), convert them to tensors, and normalize the pixel values. The dataset has been loaded with the specified transformations. The labeled data is split into a training set and a validation set. The unlabeled data is loaded using the same transformations. A ResNet50 model is defined as the teacher model, which will be used for blind distillation. The teacher model is trained on the labeled training set using a training process specific to the chosen deep learning framework. Pseudo-labels are generated for the unlabeled data by passing the images through the trained teacher model. A ResNet8 model is defined as the student model,

which will be trained using blind distillation. The student model is trained on the labeled training set and the pseudo-labeled unlabeled data, following a training process specific to the chosen deep learning framework. The student model is fine-tuned using knowledge distillation from the teacher model, utilizing a fine-tuning process specific to the chosen deep learning framework. Two identical ResNet8 models are initialized. The models are trained on their respective labeled training sets using a training process specific to the chosen deep learning framework. Predictions are generated on the unlabeled data Predictions are generated on the unlabeled data using both models. Confident predictions are selected as additional labeled examples. The labeled training data is updated with the new examples. The co-training process is repeated for a fixed number of steps. A calculate accuracy function is defined to calculate the accuracy of a given model on a given data loader. The accuracy of both models is calculated on the validation set using the calculate accuracy function. The accuracy of each model is printed to the console.

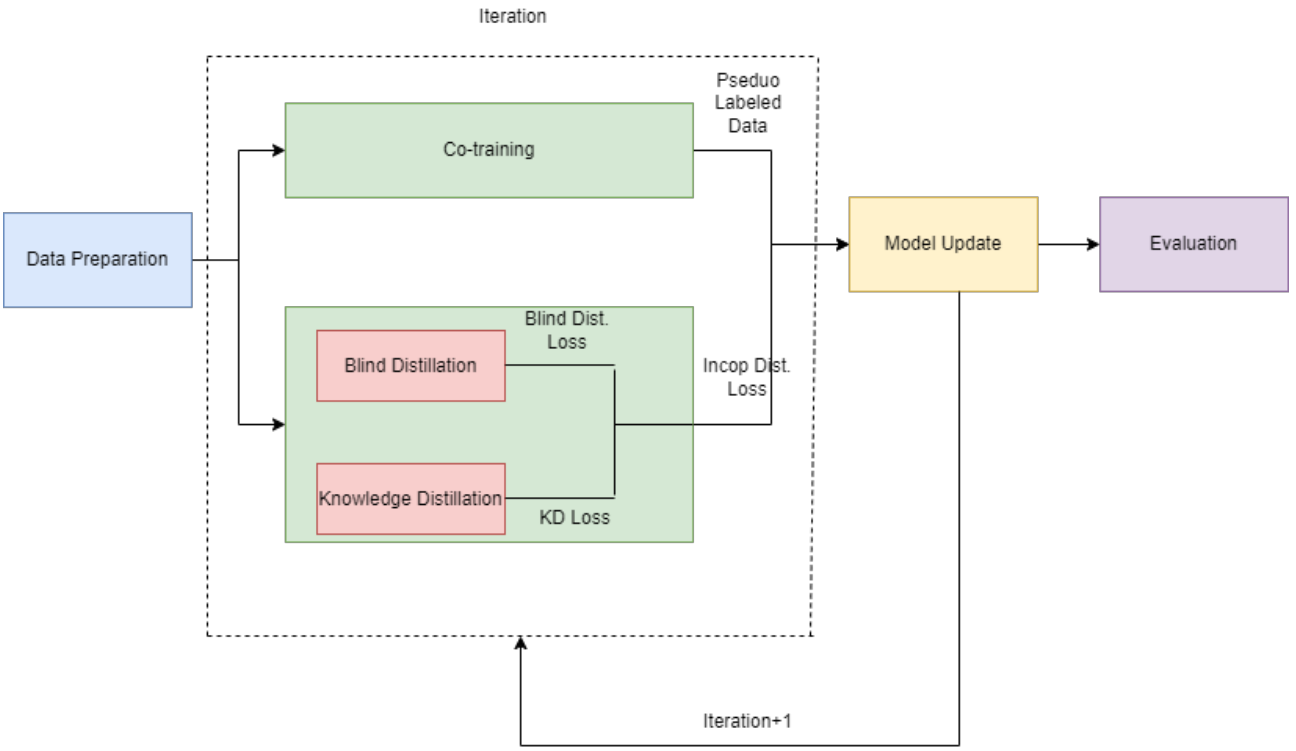

Figure 9: Blind Distillation using Co-training

CHAPTER 4

4 EVALUATION AND RESULT

This thesis talks about classifying the images using huge unlabeled data and with less labeled data. There are multiple models like Co-training, Mix-up Procedure, Augmenting both label and unlabeled, Knowledge Distillation and Blind Distillation. The datasets used are Celeb-a[25]and NIR-VIS dataset[12]. The models were compared on the basis of accuracy.

4.1 Evaluation Metrics

The Celeb-A[25] dataset has 40 different attribute labels per image. The Gender("Male" and "Female") attribute and Smiling("Smiling" and "Not Smiling") attribute has been to used to evaluate the model. For NIR-VIS dataset[12] gender attribute has been used to evaluate model. The accuracy is calculated for each model. As the models are Iterative, the first evaluation is addressed as Baseline accuracy and the next iterations accuracy are referred as 1st iteration, 2nd iteration etc. For all the experiments, a limited no. of training samples (1000 samples) are used to train a baseline classifier. The remaining samples in the training set were used as unlabeled batches. Using the proposed techniques, highly confidently classified samples from the unlabeled batch of samples (three sets available at different time stamps) are used to fine-tune the baseline classifier, trained on 1000 initial labeled samples as shown in Table 2.

Table 2: Protocol Followed for Harnessing Unlabeled Data.

Labeled samples	Celeb-A Unlabeled samples				NIR-VIS Unlabeled Batch			
	1	2	3	Test Set	1	2	3	Test Set
All Samples	0	0	0	19962	0	0	0	3744
1000 (baseline)	66333	66333	66334	19962	7868	7868	7869	3744

At each iteration of fine-tuning the model using the confidently classified samples, the performance of the model is evaluated on the test set. The validation set is used for optimal parameter selection.

4.2 Results

In this section the results of individual models are discussed,comparing the results with other existing models. All the models are trained on celeb-a[25] and NIR-VIS[12] datasets. The dataset is already divide into train, test and validation.

4.2.1 Results For Basic Co-training model

The accuracy for basic co-training model is calculated. All the models are evaluated in test set of 19867 images of celeb-a dataset and 3744 images of NIR-VIS dataset. From the table 13 accuracy of both model is computed. The Baseline accuracy of Model-1 in Celeb-A dataset is 83% and for the 3rd iteration of model 1 has the accuracy of 86.5% which the model is improved the accuracy by 3.5%. Similar to model 1, The model 2 accuracy of celeb-a dataset pumped from 83% to 85.5%.

Table 3: Gender Classification Accuracy of model 1 and model 2 for basic co-training

Iterations	Model 1			Model 2		
	Celeb-A	NIR	VIS	Celeb-A	NIR	VIS
Baseline	83	79	84	83	80.3	84.7
1st Iteration	83.9	80.5	85.3	83.4	81.2	85.5
2nd Iteration	85.7	81.8	86.7	84.2	83.5	87.2
3rd Iteration	86.5	82.4	88	85.5	84.3	87.9

Table 4: Smiling Classification Accuracy of Celeb-A for basic co-training

Iteration	Model 1	Model 2
Baseline	81	80.5
1st Iteration	82.2	81.4
2nd Iteration	84.5	82.8
3rd Iteration	85.7	84.1

Similar to Celeb-A dataset, The NIR-VIS dataset also improved, For model 1 NIR and VIS improved from 79% to 82.4% and 84% to 88 respectively. For Model 2, NIR increased from 80.3% to 84.3% and VIS increased 84.7% to 87.9%. From the table 13 the accuracy of gender classification has increased for the both model.

For Smiling or not attribute of celeb-a dataset is compared. From the table 4 baseline accuracy of model 1 is 81% to 3rd iteration accuracy is85.7% and for model 2 baseline is 80.5% to 3rd iteration is 84.1%.

For Celeb-A Dataset gender prediction and smiling classification task model 1 is ResNet50 and model 2 is Inception V3. For NIR-VIS Dataset both the models are ResNet50.

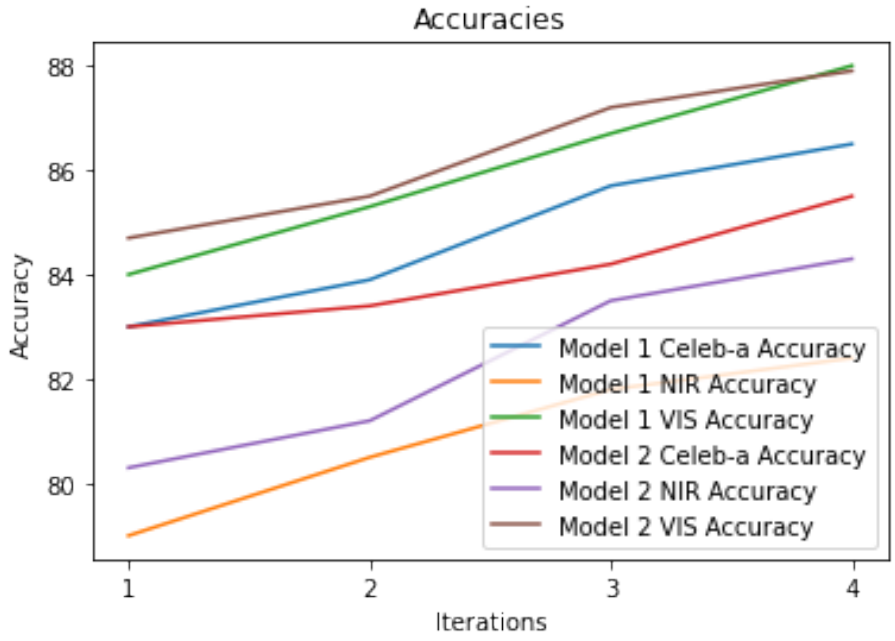

Figure 10: Visualizing the accuracies of basic-co-train

4.2.2 Results For Co-training with Mix-up Procedure

Incorporating two methods co-training and mix-up procedure has significantly improved from previous methods. From the table 5 For Celeb-A Model 1, accuracy rose from 82% to 91.2% percent, up to 9.2 percent. The baseline accuracy for the NIR and VIS datasets in model 1 is 79.4% to 86.44% increment of 7% and 83.4% to 90.34% increment of 7.3%, respectively. Similar to model 1, the accuracy of the model 2 celeb-a, NIR, and VIS is between 81.4% and 87% percent, 76.2% and 81.2% percent, and 79.8% and 84.7% percent, respectively.

For Smiling classification baseline accuracy of celeb-a dataset baseline accuracy for model 1 is 84.2%, 1st iteration has 86.3%, 2nd iteration is 88.2% and final iteration is 89.4%. Similar model 1, model 2 baseline 79.4%, 1st iteration is 81.4%, 2nd iteration is 83.5% and final iteration is 85.3%. Model 2 is increased by 7.% fo celeb-a, 6.9% and 4.8% respectively. Where this model has average of 7% increase from baseline to 3rd iteration. This model performing with the increase in accuracy.

For Celeb-A Dataset gender prediction and smiling classification task model 1 is ResNet50 and model 2 is Inception V3. For NIR-VIS Dataset both the models are ResNet50.

Table 5: Gender Classification Accuracy of model 1 and model 2 for co-training with Mix-up Procedure

Iterations	Model 1			Model 2		
	Celeb-A	NIR	VIS	Celeb-A	NIR	VIS
Baseline	82	79.4	83	81.4	76.2	79.8
1st Iteration	85.3	82.5	86.1	83.5	78.6	81.4
2nd Iteration	89.6	85.2	88.5	85.1	79.4	82.9
3rd Iteration	91.2	86.4	90.3	87	81.2	84.7

Table 6: Smiling Classification Accuracy of Celeb-A for co-training with Mix-up procedure

Iteration	Model 1	Model 2
Baseline	84.2	79.4
1st Iteration	86.3	81.4
2nd Iteration	88.2	83.5
3rd Iteration	89.4	85.3

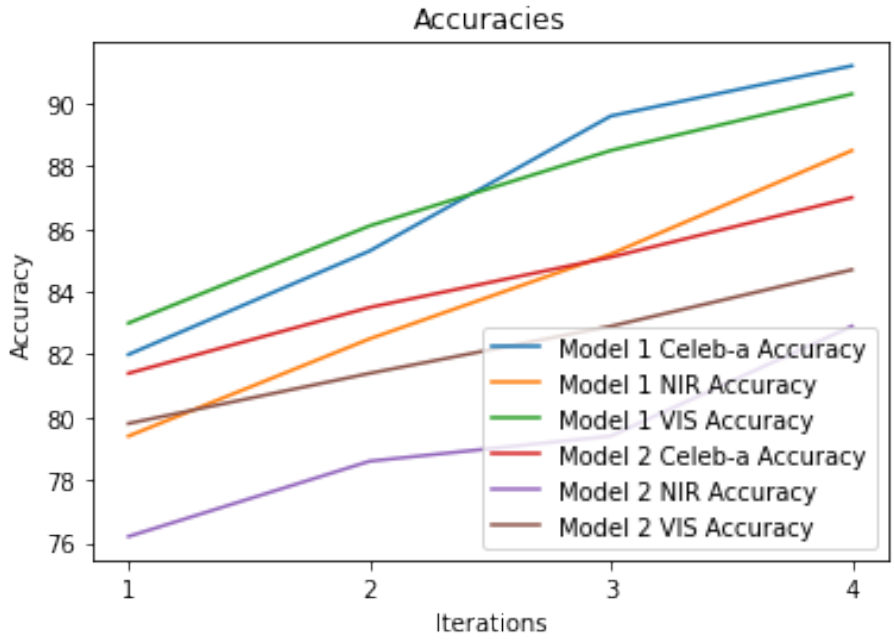

Figure 11: Visualizing the accuracies of co-training with mix-up procedure

4.2.3 Results For Co-training with Augmenting both label and unlabeled

Augmenting both label and unlabeled data and followed by the co-training technique also improved the accuracy for all datasets. from the table 7 The accuracy are improved for celeb-a model 1 from 82% to 88.8% which has increased to 6.2%. For NIR and VIS dataset in model 1 the baseline accuracy is 79.4% to 85.4% increment of 6% and 83% to 88.2% increment of 5.3% respectively. Similar to model 1, The model 2 celeb-a, NIR and VIS accuracy are 82% to 89.7%, 79.4% to 86.3% and 83% to 87.8% respectively.

The baseline accuracy for smiling classification in the Celeb-a dataset is 83.2 percent for model 1, 84.5 percent for iteration 1, 86.3 percent for iteration 2, and 87.2 percent for iteration 3. Model 2's baseline is 80 percent, the first iteration is 82.4%, the second iteration is 83.2%, and the third iteration is 84.5 percent.

For Celeb-A Dataset gender prediction and smiling classification task model 1 is ResNet50 and model 2 is Inception V3. For NIR-VIS Dataset both the models are ResNet50.

Table 7: Gender Classification Accuracy of model 1 and model 2 for co-training with Augmenting both and unlabeled

Iterations	Model 1			Model 2		
	Celeb-A	NIR	VIS	Celeb-A	NIR	VIS
Baseline	82	79.4	83	82	79.4	83
1st Iteration	84.8	80.7	84.6	84.8	81.4	84.4
2nd Iteration	87.2	83.2	86.3	87	84.2	85.9
3rd Iteration	88.8	85.4	88.2	89.7	86.3	87.8

Table 8: Smiling Classification Accuracy of Celeb-A for co-training with Augmenting both label and unlabeled

Iteration	Model 1	Model 2
Baseline	82	83.1
1st Iteration	84.2	84
2nd Iteration	86.1	85.2
3rd Iteration	87	86.4

4.2.4 Results For Knowledge Distillation with Co-training

Knowledge Distillation is used to improve the accuracy and follwed by co-training the model will perform well. The ResNet50 as a teacher model and ResNet8 as a student model to improve the performance. The accuracy of celeb-a model 1 has grown from 84% to 90.4%, with a margin of increase 6.4%. In model 1, the baseline accuracy for NIR and VIS datasets is 79.3% to 85.8% increment of 6.5% and 82.4% to 88.3% increment of 5.9%, respectively. Model 2 celeb-a, NIR, and VIS accuracy are 81.4% to 87%, 76.2% to 81.2%, and 79.8% to 84.7%, respectively, similar to model 1.

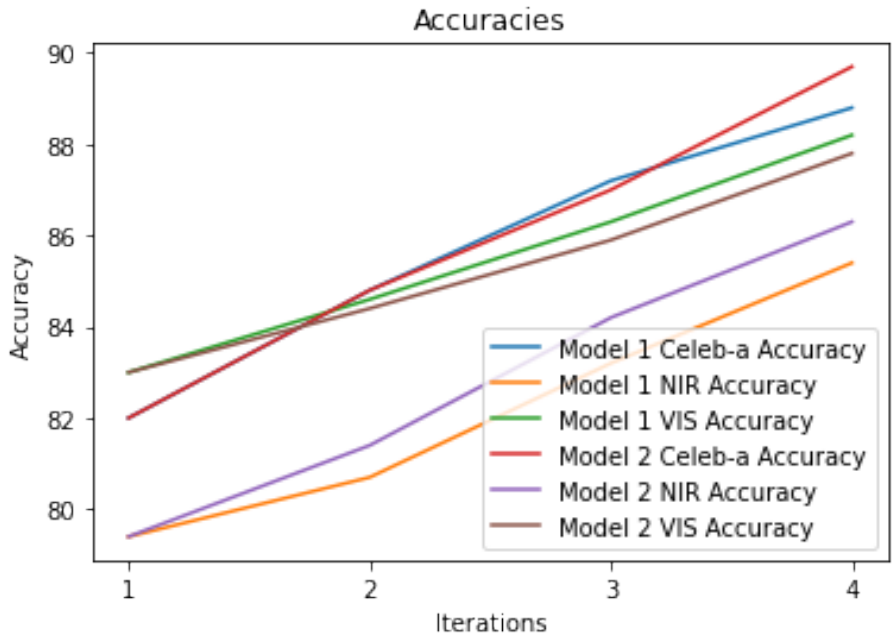

Figure 12: Visualizing the accuracies of co-training with augmenting both label and unlabeled

Table 9: Gender Classification Accuracy of model 1 and model 2 for Knowledge Distillation with Co-training

Iterations	Model 1			Model 2		
	Celeb-A	NIR	VIS	Celeb-A	NIR	VIS
Baseline	84	79.3	82.4	81.4	76.2	79.8
1st Iteration	86.1	81.6	84.7	83.5	78.6	81.4
2nd Iteration	88.8	83.3	86	85.1	79.4	82.9
3rd Iteration	90.4	85.8	88.3	87	81.2	84.7

Table 10: Smiling Classification Accuracy of Celeb-A for Knowledge Distillation with co-training

Iteration	Model 1	Model 2
Baseline	83.2	80
1st Iteration	84.5	82.4
2nd Iteration	86.3	83.2
3rd Iteration	87.2	84.5

The baseline accuracy for smiling classification in the Celeb-a dataset is 83.2 percent for model 1, 84.5 percent for iteration 1, 86.3 percent for iteration 2, and 87.2 percent for iteration 3. Model 2's baseline is 80 percent, the first iteration is 82.4%, the second iteration is 83.2%, and the third iteration is 84.5% percent.

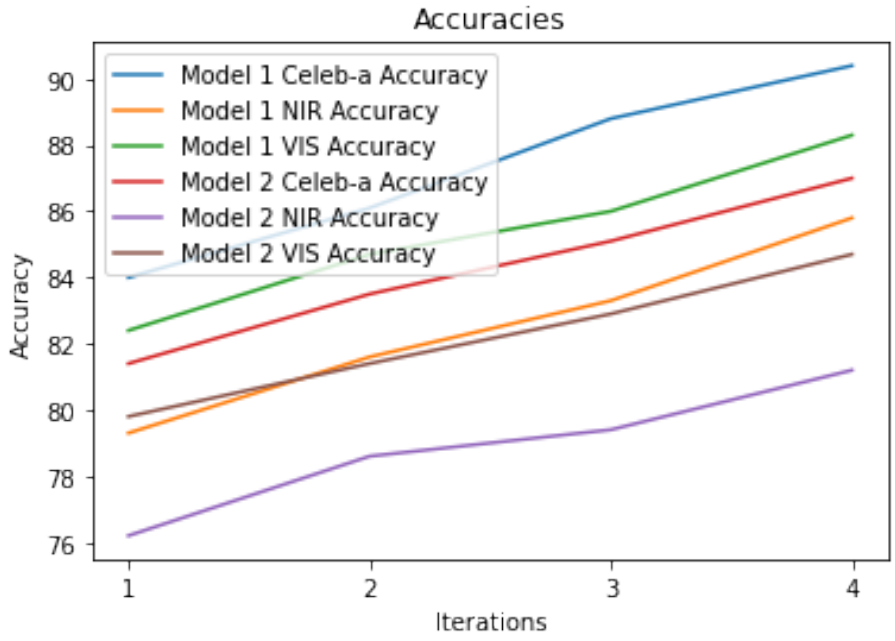

Figure 13: Visualizing the accuracies of Knowledge distillation and co-training

Table 11: Gender Classification Accuracy of model 1 and model 2 for Blind Distillation with Co-training

Iterations	Model 1			Model 2		
	Celeb-A	NIR	VIS	Celeb-A	NIR	VIS
Baseline	84	79	81	82.6	75	78.8
1st Iteration	86.2	82	83.5	85	77.4	80.5
2nd Iteration	87.5	83.4	84.2	86.1	78.6	81.8
3rd Iteration	88.3	84.8	85.6	87.5	80	82.7

Table 12: Smiling Classification Accuracy of Celeb-A for Blind Distillation with co-training

Iteration	Model 1	Model 2
Baseline	84	81
1st Iteration	85.4	82.4
2nd Iteration	86	84.1
3rd Iteration	87.5	85.7

4.2.5 Results For Blind Distillation with Co-training

Blind Distillation is similar to knowledge distillation method which improves the accuracy of the given model. from the table **??** the baseline accuracy of model 1 teacher model ResNet50 celeb-a dataset is 84%, 1st iteration is 86.2%, 2nd iteration is 87.5% and 3rd iteration is 88.3% which gives hike of 4.3%. Model 1 NIR started with 79% and the 3rd iteration is 84.8%, VIS baseline is 81% and final iteration is 84.8%. Likewise, Model 2 student model ResNet8 celeb-a dataset started with 82.6% followed by 85%, 86.1% and 87.5% increase of 4.9%. For NIR-VIS dataset baseline of 75% and 78.8% respectively and they are rose upto 80% and 82.7% repectively.

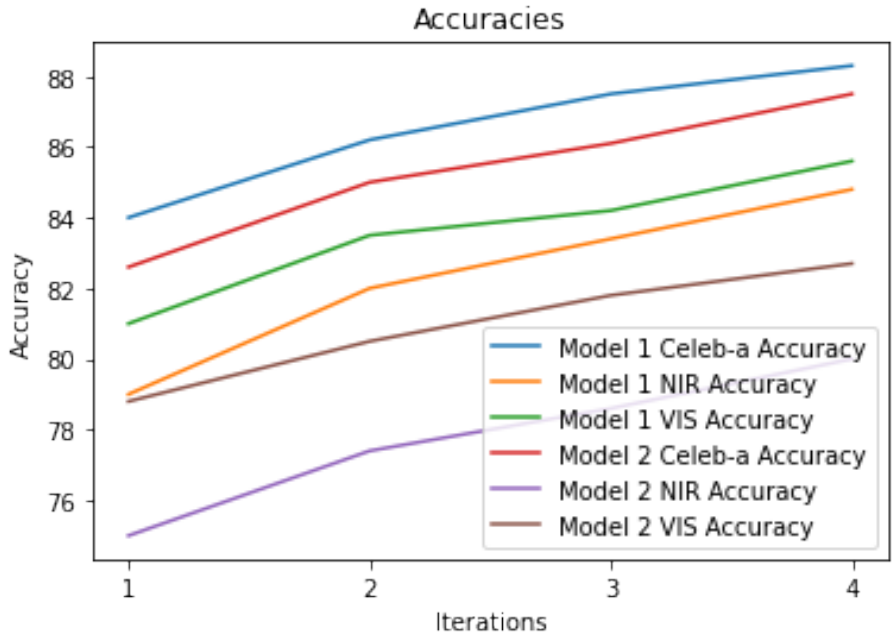

Figure 14: Visualizing the accuracies of Blind Distillation with co-training

Table 13: Gender Classification Accuracy of model 1 and model 2 for all techniques

Techniques	Model 1			Model 2		
	Celeb-A	NIR	VIS	Celeb-A	NIR	VIS
Basic Co-Train	83 ± 86.5	79 ± 82.5	84 ± 88	83 ± 85.5	80.3 ± 84.3	84.7 ± 87.9
Co-train with Mix-up	82 ± 91.2	79.4 ± 86.4	83 ± 90.3	81.4 ± 87	76.2 ± 81.2	79.8 ± 84.7
Co-train with Aug	82 ± 88.8	79.4 ± 85.4	83 ± 88.2	82 ± 89.7	79.4 ± 86.3	83 ± 87.8
KD with Co-train	84 ± 90.4	79.3 ± 85.8	82.4 ± 88.3	81.4 ± 87	76.2 ± 81.2	79.8 ± 84.7
BD with Co-train	84 ± 88.3	79 ± 84.8	81 ± 85.6	82.6 ± 87.5	75 ± 80	78.8 ± 82.7

Table 14: Smiling Classification Accuracy of Celeb-A for all techniques

Techniques	Model 1	Model 2
Basic Co-Train	81 ± 85.6	80.5 ± 84.1
Co-train with Mix-up	84.2 ± 89.4	79.4 ± 85.3
Co-train with Aug	82 ± 87	83.1 ± 86.4
KD with Co-train	83.2 ± 87.2	80 ± 84.5
BD with Co-train	84 ± 87.5	81 ± 85.7

Smiling attribute of Celeb-a dataset, Model 1 baseline accuracy is 84% and by updating the samples the 1st iteration is 85.4%, 2nd iteration is 86% and 3rd iteration is 87.5%. computing the increase accuracy of 3.5%. Same as model 1, model 2 started with 81%, 1st iteration is 82.4%, 2nd iteration 84.1% and 3rd iteration is 85.7%. Model 2 incrsead performance by 4.7%. Both the models performance are increasing.

4.2.6 Comparison of state of art techniques

It compares the accuracy improvements of different models using various techniques, including co-training, mix-up procedure, knowledge distillation, and blind distillation. The comparisons are based on evaluations conducted on the Celeb-A dataset and the NIR-VIS dataset.

In the co-training models, both Model 1 and Model 2 demonstrated improvements in accuracy on the Celeb-A dataset. Model 1's accuracy increased from a baseline of 83% to 86.5% in the 3rd iteration, representing a 3.5% improvement. Similarly, Model 2 showed an accuracy improvement from 83% to 85.5%. The NIR and VIS datasets also witnessed accuracy enhancements, with Model 1 showing improvements from 79% to 82.4% for NIR and from 84% to 88% for VIS. Model 2 exhibited increases from 80.3% to 84.3% for NIR and from 84.7% to 87.9% for VIS. Additionally, both models demonstrated increased accuracy in gender classification.

For the attribute of smiling or not in the Celeb-A dataset, both models showed accuracy improvements. Model 1 increased its accuracy from a baseline of 81% to 85.7% in the 3rd iteration, while Model 2 improved from a baseline of 80.5% to 84.1%. Incorporating co-training and the mix-up procedure led to significant accuracy improvements in both models. In Model 1, the accuracy rose from 82% to 91.2%, representing a 9.2% improvement. The NIR and VIS datasets in Model 1 also experienced accuracy increases of 7% and 7.3% respectively. Model 2 demonstrated similar improvements, with accuracy increases ranging between 7% and 7.3% across multiple datasets and attributes.

The knowledge distillation method, where a ResNet50 teacher model guides a ResNet8 student model, also resulted in accuracy improvements. For Model 1, the accuracy of the Celeb-A dataset increased from 84% to 90.4%, with NIR and VIS accuracies improving by 6.5% and 5.9% respec-

tively. Model 2 exhibited accuracy improvements of 4.9%, 5.0%, and 4.9% for Celeb-A, NIR, and VIS datasets respectively. The smiling classification also showed accuracy improvements for both models.

Blind distillation, similar to knowledge distillation, was effective in improving model accuracy. Model 1, using the blind distillation method, increased its accuracy from a baseline of 84% to 88.3% in the Celeb-A dataset. Similar accuracy improvements were observed for NIR and VIS datasets in Model 1. Model 2 also demonstrated accuracy improvements across datasets and attributes.

Overall, the different techniques employed in these models, including co-training, mix-up procedure, knowledge distillation, and blind distillation, resulted in consistent accuracy improvements across various datasets and attributes. These techniques proved to be effective in enhancing model performance, providing higher accuracy in image classification tasks.

CHAPTER 5

5 CONCLUSION AND FURTHER WORK

The comparison of different models using various techniques highlights the effectiveness of co-training, mix-up procedure, knowledge distillation, and blind distillation in improving accuracy in image classification tasks. The models consistently showed accuracy improvements across the Celeb-A dataset, NIR-VIS dataset. Overall, the findings from the comparisons suggest that augmenting label and unlabeled data, incorporating co-training, and utilizing techniques such as mix-up procedure, knowledge distillation, and blind distillation can significantly enhance the accuracy of image classification models. These techniques provide promising avenues for improving model performance and achieving higher accuracy rates in real-world applications. Based on the comparisons and conclusions drawn from the above analysis, there are several potential directions for further work in the field of image classification. Future research could investigate the synergistic effects of combining multiple approaches and identify the optimal combination for specific datasets and attributes. Extending the analysis to other computer vision tasks, such as object detection, semantic segmentation, or instance segmentation, could reveal the effectiveness of the proposed techniques in various domains. It would be interesting to explore how the models perform in tasks beyond image classification.

REFERENCES

REFERENCES

[1] A. K. Jain, S. C. Dass, K. Nandakumar *et al.*, "Soft biometric traits for personal recognition systems," in *ICBA*, vol. 3072. Springer, 2004, pp. 731–738.

[2] A. V. Nadimpalli, N. Reddy, S. Ramachandran, and A. Rattani, "Harnessing unlabeled data to improve generalization of biometric gender and age classifiers," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2021, pp. 1–7.

[3] X. J. Zhu, "Semi-supervised learning literature survey," 2005.

[4] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 92–100.

[5] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[6] G. Menghani and S. Ravi, "Learning from a teacher using unlabeled data," *arXiv preprint arXiv:1911.05275*, 2019.

[7] T. Miyato, S.-i. Maeda, M. Koyama, and S. Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1979–1993, 2018.

[8] 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.

[9] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. X. Han, Z. Zhao, and B. Li, "Age group and gender estimation in the wild with deep ror architecture," *IEEE Access*, vol. 5, pp. 22 492–22 503, 2017.

[10] N. Narang and T. Bourlai, "Gender and ethnicity classification using deep learning in heterogeneous face recognition," in *2016 International Conference on biometrics (ICB)*. IEEE, 2016, pp. 1–8.

[11] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 3730–3738.

[12] J. Bernhard, J. Barr, K. W. Bowyer, and P. Flynn, "Near-ir to visible light face matching: Effectiveness of pre-processing options for commercial matchers," in *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2015, pp. 1–8.

[13] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015.

[14] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using em," *Machine learning*, vol. 39, pp. 103–134, 2000.

[15] F. Ma, D. Meng, X. Dong, and Y. Yang, "Self-paced multi-view co-training," *Journal of Machine Learning Research*, 2020.

[16] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," in *Proceedings of the european conference on computer vision (eccv)*, 2018, pp. 135–152.

[17] Y. Tokozume, Y. Ushiku, and T. Harada, "Between-class learning for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 5486–5494.

[18] T. Miyato, A. M. Dai, and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," *arXiv preprint arXiv:1605.07725*, 2016.

[19] K. Saito, Y. Ushiku, and T. Harada, "Asymmetric tri-training for unsupervised domain adaptation," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2988–2997.

[20] A. Romero, N. Ballas, S. E. Kahou, A. Chassang, C. Gatta, and Y. Bengio, "Fitnets: Hints for thin deep nets," *arXiv preprint arXiv:1412.6550*, 2014.

[21] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *arXiv preprint arXiv:1612.03928*, 2016.

[22] G. Chen, W. Choi, X. Yu, T. Han, and M. Chandraker, "Learning efficient object detection models with knowledge distillation," *Advances in neural information processing systems*, vol. 30, 2017.

[23] B. Heo, M. Lee, S. Yun, and J. Y. Choi, "Knowledge transfer via distillation of activation boundaries formed by hidden neurons," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 3779–3787.

[24] T. Furlanello, Z. Lipton, M. Tschannen, L. Itti, and A. Anandkumar, "Born again neural networks," in *International Conference on Machine Learning*. PMLR, 2018, pp. 1607–1616.

[25] S. Yang, P. Luo, C.-C. Loy, and X. Tang, "From facial parts responses to face detection: A deep learning approach," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 3676–3684.