CHAPTER-2

Literature Review

2.1 Introduction

A literature review is a continuous learning process which has been carried out throughout the course of research. The major objective of this chapter is to gain enough background knowledge on the facial retouching. The literature review for this study is broken down into five distinct sections. Table 2.1 provides section-by-section information on these phases, along with their goals and conclusions.

TABLE 2.1: Different phases of Literature Review

Phase	Area of Literature Review	Objective	Outcomes
1	Retouching	To understand the retouching on face images.	Different types of retouching over real face images are studied.
2	Related Work	To find capable and stable algorithms for the retouching classification.	Different deep learning algorithms and their limitations are studied.
3	Transfer Learning Approach	To find CNN & TL models used for specifically facial retouching task.	Brief awareness of basic structure of CNN is gained.
4	Pre-trained Models	To select the CNN model for classification task.	VGG16 and ResNet50 models are chosen and studied in detailed.
5	Optimizer for classification task	To find out the most versatile Optimizer considered in the literature for improving classification accuracy .	Adam and RMSprop optimizers give better improvement compared to others.

Systematic literature survey throughout research work is useful to find the following information:

- Different types of retouching techniques applied over face images.
- Different algorithm and deep learning approach used for retouching classification.
- Different types of optimizers used for classification task, their merits, and demerits.
- Different factors affecting the performance of a classification accuracy of the TL model.

2.2 Overview of facial retouching and detection

Optimization Facial retouching is a common practice in photography and image editing, often used to enhance the appearance of a person's face in various ways, such as smoothing skin, removing blemishes, reducing wrinkles, and improving overall aesthetics[4]. While retouching can be used for artistic and cosmetic purposes, it can also lead to unrealistic and misleading representations of individuals, which has raised concerns about the need for detection and disclosure of retouching in various contexts, including advertising, social media, and journalism. In some places, there have been calls for legislation or industry guidelines requiring the disclosure of retouching in advertising and editorial content. France, for instance, introduced a law that requires digitally altered images to carry a "retouched photo" label[5]. Detecting facial retouching in images is challenging but not impossible. Several methods can be used to identify retouched images like, human experts, such as photo editors and forensic analysts, can visually inspect images for signs of retouching. Software tools and algorithms can analyse images to detect unusual patterns, such as overly smooth skin, inconsistent lighting, or altered facial proportions, which may indicate retouching. Content creators, including photographers, makeup artists, and image editors, have a responsibility to maintain ethical standards in their work and disclose when significant retouching has been applied. Promoting media literacy and educating the public about the potential for retouching in images can help individuals become more critical consumers of visual content. Efforts to regulate and disclose retouching are also ongoing in some regions, emphasizing the importance of transparency and ethics in visual media.[6].

2.3 Related work on facial retouching classification

Retouching is a doctoring technique to be done over any digital images. This attack can be active or passive. The branches of forgery attacks are again different based on the altering done on images. The performance of the forgery detection is based on the dataset used and the software applications used for implementation[7][8]. Reference [4] developed a perceptual matric learned on support vector regression (SVR) to estimate the map the between user rating and summative statics of the retouched images (geometric and photometric alteration).The real and retouched images of total 468 images are collected from different on-line sources.

The ND-IIITD dataset were developed consists of 2600 un-retouched and 2275 retouched images. the database contains male and female face images. Research work carried out in Reference [9] used unsupervised and supervised deep learning algorithm for detecting the retouched and real images of the ND-IIITD database. The same approach is carried out for Celebrities database consists of 165 objects (330 real plus retouched) from on-line sources. The overall accuracy achieved was 90.9% and 96.8% for ND-IITD and Celebrities dataset respectively using unsupervised DBM. The overall accuracy achieved was 93.9% and 98.7% for ND-IITD and Celebrities dataset respectively using supervised DBM.

In 2017, the algorithm was proposed which uses semi supervised auto encoders to report the retouching accuracy on the Multi-Demographic Retouched Faces (MDRF) dataset[10]. MDRF dataset is introduced by the author with subjects from three different ethnicities and forgery from two tools is applied.

Moreover, besides using photo editing tools, the Generative Adversial Network(GAN) generated retouched images are widely used to train the deep learning models. Reference [11] proposed a CNN approach to detect and classify retouched images of ND-IIITD retouched faces dataset and CelebA dataset. The real images of CelebA dataset is used to generate the retouched images using StarGAN. 99.70% and 99.42% accuracy is achieved which is ~6% higher compared to [9] using Thresholding and SVM classifier.

Studies demonstrate that when a makeup is applied, face recognition systems doesn’t work well. The publically available makeup dataset are YMU(YouTube Makeup Dataset)[12], MID(Makeup in the wild dataset)[13] and FCD(facial Cosmetic Dataset)[14]. The research described in [15] was able to extract a features vector that accurately depicted the input face's shape and texture. Following feature extraction, two different classifiers—namely, SVM and Alligator—are used for comparison. 99.30% overall accuracy is achieved for inter database classification.

Plastic surgery is another type of forgery class which offers adverse effect on face recognition task. Reference [16] offer an experimental study to quantitatively assess face recognition algorithms' performance on a database of people who have undergone both local and global plastic surgery. The research demonstrates that the algorithms PCA, FDA, GF, LFA, LBP, and GNN are unable to successfully offset the variances brought on by the plastic surgery treatments where overall accuracy achieved is ~34% maximum.

As variety of photo editing tools are available freely and easily, photo retouching can be done even any layman with ease. The evaluation of 32 beauty apps were conducted in [17]. A database of 800 enhanced face photos is created using five apps namely Airbrush, Instabeauty, Fotorush, Polarr and Youcam perfect based on this evaluation. A commercial face recognition system is used to compare biometric performance before and after retouching. The analysis of photo response non-uniformity (PRNU), which forms the basis of a retouching detection system, is provided and the approach achieved the avg detection error rate 13.7%. In 2022, reference [18] introduce IPDCN2, an enhanced patch-based deep convolutional neural network designed to differentiate between original and retouched facial images. the extracted relevant patches from the input image using 68 facial landmarks are pre-processed and utilize residual learning to maximize information flow throughout the neural network.

2.4 Transfer learning approach in image classification

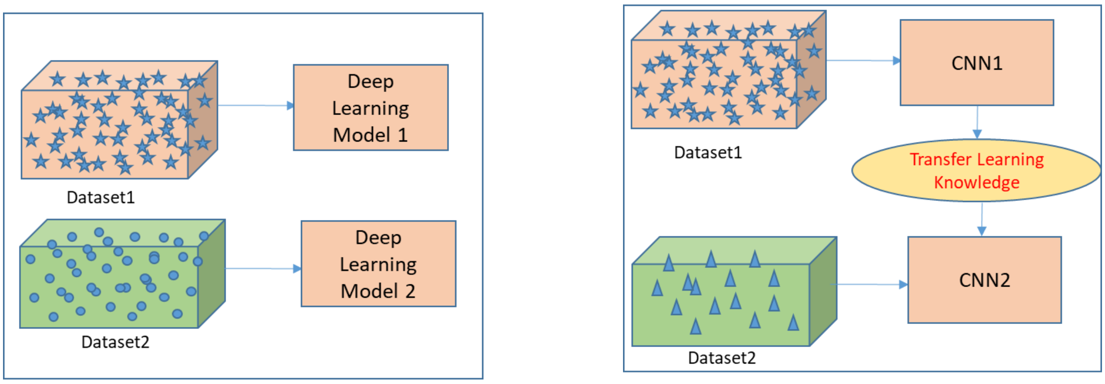
Transfer Learning can be used in variety of fields like medical, weather reporting, forecasting, road map detection, image retouching to classify the deceases, or cancer or tumours, sky conditions, map detection and to detect forgery on images, etc. As compared to ML & DL approach, TL(transfer learning) approach are faster and trained more accurately than other traditional methods like manual grading and other machine vision techniques or other classifiers[19]. there are several challenges, when using DL(Deep Learning) model to detect retouching on facial images. Some of them are listed here,

1. Need of large facial dataset containing real and retouched face images

2. Proper labelled data of images

3.Large amount of images for training the model

4. DL models are prone to overfitting too which leads to give biased output.



(a) (b)

FIGURE 2.1: Difference between DL approach(a) and TL Approach(b)

Using, TL, all these challenges are overcome and optimal detection accuracy can be achieved. TL offers following advantages in ML and DL tasks[20],

1. Reduced Training Time: Transfer learning allows you to leverage pre-trained models that have been trained on large datasets. By using a pre-trained model as a starting point, you can save a significant amount of time and computational resources that would otherwise be required to train a model from scratch.
2. Lower Data Requirements: Training deep learning models often requires large amounts of labelled data. Transfer learning enables you to overcome this challenge by using the knowledge gained from a source task (where data may be abundant) to improve the performance on a target task with limited data, as depicted in fig. 2.1. This is particularly beneficial in scenarios where collecting large amounts of labeled data is difficult or expensive.
3. Improved Generalization: Transfer learning helps improve the generalization capabilities of a model. Pre-trained models have learned useful features from a diverse range of data, which can be transferred to a new task. This transfer of knowledge allows the model to extract relevant features from the data more effectively, even when the target task has different characteristics or a smaller dataset.
4. Avoiding the "Cold Start" Problem: When starting a new machine learning project, especially with limited data, it can be challenging to train an accurate model from scratch. Transfer learning helps overcome the "cold start" problem by providing a well-

initialized model that has already learned useful representations from a different but related task.

5. Effective in Domain Adaptation: Transfer learning is highly useful in domain adaptation scenarios, where the distribution of the source data differs from the target data. By using a pre-trained model on a source domain and fine-tuning it on the target domain, transfer learning enables the model to adapt and perform well in the target domain.
6. Useful for Small-Scale Deployment: In resource-constrained environments, such as edge devices or mobile applications, transfer learning allows you to deploy efficient models that consume less computational power and memory. By leveraging pre-trained models and fine-tuning them on specific tasks, you can achieve good performance with fewer resources.

2.4.1 Previous studies on VGG16 and ResNet50 models

Several pre-trained networks like VGG16, Inception, Xception, ResNet50, MobileNet, AlexNet, DenseNet have gained prominence in the field of deep learning and computer vision. These pre-trained networks differ in their architectural designs and innovations. VGG16 and AlexNet are known for their simplicity, while ResNet50, InceptionV3, and Xception incorporate specialized features like residual connections, multiple kernel sizes, and depthwise separable convolutions to improve performance and efficiency. Models like MobileNet and DenseNet are tailored for resource-constrained environments, offering reduced model size and computational requirements. Each of these networks addresses specific challenges and trade-offs in the realm of deep learning for computer vision.

the VGG16 model and its deeper variants (VGG19) along with their application discusses in large-scale image recognition tasks[21]. It has been highly influential in the field of computer vision and deep learning. Researchers and practitioners interested in using VGG16 for image classification tasks often refer to this paper for insights into the architecture and training methods. Traditionally, data mining algorithms and machine learning algorithms are engineered to approach the problems in isolation. Transfer learning method is used by reusing a pre-trained model knowledge for classifying the cat and dog images task[22]. The work demonstrated the validation accuracy is increased by ~15% with fine tuning of VGG16 and applying data augmentation to classify images. The significance of early detection and accurate diagnosis of skin cancer and the potential of deep learning in medical image analysis, particularly for skin cancer diagnosis is highlighted in reference[23] . It introduces the use of the pre-trained VGG16 architecture for skin cancer image classification, demonstrating promising results, with a classification accuracy of 84.242% when utilizing the YCbCr color scale, and it further explores the performance of VGG16 with images of different color scales, as well as the analysis of feature parameters from various layers for classification.

The ResNet architecture and the benefits of residual connections are discusses in training very deep neural networks[21]. It includes the ResNet50 variant, along with other ResNet models, for image classification tasks. Transfer learning was employed with three pre-trained models (MobileNet V2, ResNet50, and VGG19) for a classification task on a previously unseen dataset[24]. VGG19 demonstrated the highest classification accuracy and f1-score at 95%, albeit with a longer execution time, while ResNet50 and MobileNet V2 achieved accuracies of 92% and 93% respectively. For accurate classification of rock photos, machine learning methods are becoming more and more common. In reference [25], the data sets are divided based on rock images captured under a white light source using the Resnet 50 neural network model. The intelligent classification of rocks is carried out by continuously modifying the parameters of each layer. The final validation accuracy was reached to 94.12%.

Based on the extensive literature review, it is evident that VGG16 and ResNet50 emerge as robust and promising choices for image classification tasks, owing to their established effectiveness and performance in various computer vision applications. Hence, for facial retouching classification and detection, the transfer learning pre-trained models VGG16 and ResNet50 are used for this research.

2.5 Optimizers Selection for classification

Deep learning optimizers are an essential component in the field of computer vision since they guarantee that the training process yields the best results. By continuously changing the model's parameters, the optimizer's job is to reduce the loss, which measures the difference between expected and actual values. The effectiveness of the training, its speed and precision, and the results themselves can all be considerably impacted by selecting the correct optimizer. Therefore, optimizers are essential in computer vision applications of deep learning. The state-of-the-art optimizers used for classifications are Gradient Descent, Stochastic Gradient Descent, Stochastic Gradient descent with momentum, Mini-Batch Gradient Descent, Adagrad, RMSProp, AdaDelta, and Adam. In this research, the Adam and RMSprop are selected for their advantages over others for classification problem[26].

2.5.1 RMSprop (Root Mean Squared Propagation)

It was proposed by Geoffrey Hinton, the father of back-propagation. The algorithm's main goal is to shorten the number of function evaluations necessary to attain the local minimum, therefore quickening the optimisation process. The algorithm divides the gradient by the square root of the mean square and keeps a moving average of the squared gradients for each weight[27]. RMSprop was created for mini batch learning as a stochastic technique. It does not have a specific loss function associated with it. Instead, RMSprop is an optimization algorithm used during the training process of ML or DL models, and it works in conjunction with a specific loss function used for the given task. Typically, a common loss function, such as mean squared error (MSE) for regression or categorical cross-entropy for classification, is used in conjunction with RMSprop to update model weights during training. The loss and weight are formulating as follows:

Loss,

$$E = \frac{1}{2} * \sum_{i=1}^{t} (y_i - \hat{y_i})^2 \qquad (2.1)$$

Where,

y_i & y_i are true and predicted values respectively

Weight,

$$w_t = w_{t-1} - \tilde{\eta}_t * \frac{\partial E}{\partial w_{t-1}} \qquad (2.2)$$

Where,

Learning rate hyper parameter, $\tilde{\eta_t} = \frac{\eta}{\sqrt{M_{dw_t} + \epsilon}}$ (2.3)

Exponential moving Average, $M_{dw_t} = \beta * \gamma_{t-1} + (1 - \beta) \frac{\partial E}{\partial w_{t-1}}$ (2.4)

ϵ = Small positive number to avoid zero division

2.5.2 Adam (Adaptive Momentum Estimation)

The Adam optimizer, short for Adaptive Moment Estimation, is a popular optimization algorithm used in training deep learning models. Adam additionally stores an exponentially decaying average of past gradients, analogous to momentum, in addition to keeping an exponentially decaying average of past squared gradients, like Adadelta and RMSprop[26]. The mathematical expression to compute the weight is as follows:

$$w_t = w_{t-1} - \left(\frac{\eta}{M_{dw_t}}\right) * V_{dw_t} \tag{2.5}$$

Where,

$$V_{dw_t} = \beta_1 * V_{dw_{t-1}} + (1 - \beta_1) * \frac{\partial E}{\partial w_{t-1}} \tag{2.6}$$

$$M_{dw_t} = \beta_2 * \gamma_{t-1} + (1 - \beta_2)\frac{\partial E}{\partial w_{t-1}} \tag{2.7}$$

Here, M_{dw_t} and $V_{dw_t}$are the first and second moment estimates of the gradients, respectively. $\frac{\partial E}{\partial w_{t-1}}$ represents the gradient of the loss with respect to the model parameters. $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are hyper parameters that control the exponential decay rates of the moment estimates. To get better result at the initial time stamps, the bias correct is introduced. Accordingly, weight is updated. The formula is as follows,

$$V_{dw_t}^{corrected} = \frac{V_{dw_t}}{\left(1 - {\beta_1}^t\right)} \tag{2.8}$$

$$M_{dw_t}^{corrected} = \frac{M_{dw_t}}{\left(1 - {\beta_2}^t\right)} \tag{2.9}$$

$$w_t = w_{t-1} - \left(\eta \middle/ \sqrt{M_{dw_t}^{corrected} + \epsilon}\right) * V_{dw_t}^{corrected} \tag{2.10}$$

2.6 **Loss Function**

To calculate the error for binary classification a binary cross-entropy loss function is used, which measures the dissimilarity between the predicted probabilities and the true binary labels (real or fake face). The binary cross-entropy loss for is calculated by the following equation,

$$L_{ce} = -\sum_{j=1}^{b}\sum_{k=1}^{n} R_{jk}\ \log\left(Pr_{jk}\right) + \left(1 - R_{jk}\right)\ \log\left(1 - Pr_{jk}\right) \tag{2.11}$$

Here, L_{ce} is Binary Cross Entropy Loss, R_{jk} is True binary label of a face image (either 0 or 1), Pr_{jk} is the predicted probability output by the sigmoid function (between o to 1), b is batch size , n is no. of pixels in the image.

The sigmoid function is commonly used in binary classification problems to generate the output of a neural network model as a probability score between 0 and 1. the generated error output for a retouching classification (real =1 and fake = 0) is calculated using the sigmoid function, by the following equation,

$$Pr_j = \frac{1}{1 - e^{-\theta_j}} \tag{2.12}$$

$$\theta_j = \sum\nolimits_{k=1} w_{jk} h_k \tag{2.13}$$

Here, Pr_j is the output of the sigmoid function, θ_j is the input to the sigmoid function, which is typically the weighted sum of the model's input features and parameters.

2.7 **Summary**

In the initial phase of the literature review, an extensive collection of research papers, journals, and articles focused on a specific type of facial manipulation, referred to as "retouching," was examined. Researchers have explored various machine learning and deep learning algorithms, deploying different classifiers for image analysis. While the literature extensively covers the application of deep learning algorithms and Convolutional Neural

Networks (CNN) across diverse real-world contexts, no prior research has been identified that employs a Transfer Learning (TL) approach for the classification of genuine and manipulated facial images. Given the vast knowledge available within pre-trained TL models like VGG16 and ResNet50, their potential advantages in addressing the facial retouching classification challenge are investigated. The literature also includes documented research on the application of TL models, VGG16 and ResNet50, for other classification tasks. To harness the benefits of TL models, fine-tuning is applied to adapt VGG16 and ResNet50 models specifically for the task of facial image classification. Moreover, the study reveals the utilization of two optimization techniques to enhance model performance. Based on the comprehensive literature survey, the Adam and RMSprop optimizers have been carefully selected and implemented for further analysis, demonstrating the synthesis of existing research in the pursuit of improved facial image classification.