Chapter 5

Occlusion Handling in Image-based Person Re-Identification

In the previous two chapters, we develop person re-identification approaches that are suitable for application in only constrained scenarios devoid of any occlusion. The approach discussed in Chapter 3 averages normalized silhouettes before carrying out feature extraction in the spatial domain. This method cannot be applied in occluded situations since the presence of occlusion is likely to degrade the quality of image frames which will also drastically affect the effectiveness of the spatial-domain features. Similarly, the temporal-domain features derived in Chapter 4 using an ensemble of three time-series Neural Network models will be effective for re-identification only if the frames of the input sequence are devoid of occlusion.

As already explained in Chapter 2, a few recent work [153–156] attempt to solve the problem of re-identification in the presence of occlusion and illumination variation. Among these, Zhuo et al. in [153] considered the separate occluded/unoccluded classification and person re-identification tasks as a combined task, and does not provide any separate module for carrying out the occlusion reconstruction. In another work [154], He et al. proposed a background-foreground classifier to remove the background clutters.

Gao et al. [155] proposed a pose-guided Visible Part Matching (PVPM) model to learn attentions with discriminative part features. In [156] a cross-graph embedded-alignment (CGEA) layer has been used to embed topology information to local features and predict similarity scores for matching. Although these approaches have shown satisfactory results, it appears that the effectiveness of re-identification can be improved if two separate dedicated networks are used for occlusion reconstruction and re-identification, which we study in this chapter.

In this chapter, we focus on real-life surveillance scenarios where person images can get partially occluded due to the presence of other objects, such as tree leaves, cars, persons, etc. The presence of occlusion makes capturing of a clean walking sequence of individuals by surveillance cameras very difficult. Developing an effective algorithm to reconstruct the occluded/missing frames is expected to improve the re-identification accuracy. Here, we will specifically focus on developing effective algorithms for occlusion reconstruction in image-based person re-identification by exploiting the spatial information from the images using robust Deep Neural Network-based generators. Once the occluded frames in a sequence are reconstructed, person re-identification is carried out using any baseline network. The overall framework of the re-identification approach is given in Figure 5.1.

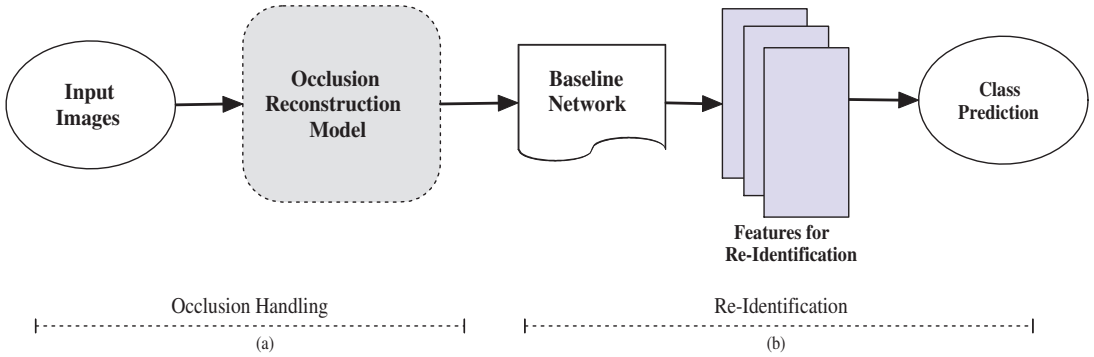

Figure 5.1: Overall framework of the re-identification approach: (a) Occlusion reconstruction, and (b) Re-identification

5.1 Synthetic Occlusion Generation for Training Deep Neural Network Models

As mentioned above, we use Deep Neural Networks to predict the unoccluded versions of input occluded frames. Since training any Machine/Deep Learning model requires the availability of ground truth, we need to prepare an extensive data set of occluded image frames along with their corresponding unoccluded counterparts and, henceforth, train a neural model to map occluded images to their corresponding unoccluded versions. Generation of this training data can be done by considering any unoccluded data set and applying synthetic occlusion at random positions within the frames present in this data set. The objects used for creating synthetic occlusion are everyday objects such as table, stool, tree, lamp, dog, etc. For each unoccluded image frame, we randomly decide the maximum percentage of pixels to occlude and superimpose certain occluding objects (randomly scaled) at randomly selected positions in the frame such that the number of pixels altered does not exceed the decided maximum percentage of occlusion. The percentage of occlusion can be defined as:

$$Percentage\ of\ Occlusion = \frac{No.\ of\ occluded\ pixels\ in\ a\ frame}{Total\ no.\ of\ pixels\ in\ a\ frame} * 100. \tag{5.1}$$

During implementation, the maximum percentage of occlusion in each frame is varied between *0* to *50*%, and a synthetically occluded re-identification data set is constructed by corrupting each frame with a certain percentage of synthetic occlusion. A few examples of corrupting images with synthetic occlusion are shown in Figure 5.2. In this figure, the first image is an original image from the *CUHK_03* data, and the next three images are synthetically occluded versions of the first image. The yellow borders shown in the next three images only highlight the synthetically occluded regions, and these borders are not present in the generated synthetically occluded data.

Figure 5.2: An original frame from the *CUHK_03* data and the corresponding occluded frames generated by adding varying degrees of synthetic occlusion

5.2 Occlusion Handling in Image Frames

To reconstruct the occluded frames, we study the applicability of two different Neural Network-based models, namely, *OHGAN* and *Autoencoder* both of which exploit the spatial information from the image frames. These two approaches are discussed in Sections 5.2.1 and 5.2.2, respectively.

5.2.1 *OHGAN*-based Reconstruction

An insight view of *OHGAN* is shown in Figure 5.3. With reference to this figure, the generator of the *OHGAN* learns a non-linear projection function to map an input occluded image to the corresponding occlusion-free image by rendering the occluded pixels with appropriate colors. This generator is based on the popular U-Net [157] architecture and has a layered configuration consisting of convolution and de-convolution layers (see Table 5.1). The convolution layers encode the basic spatial information from the input image while eliminating the occlusion and provide feature representations to the de-convolution layers, which then restore the input image by decoding the feature vectors from the encoded samples. The discriminator of the *OHGAN* is a Siamese

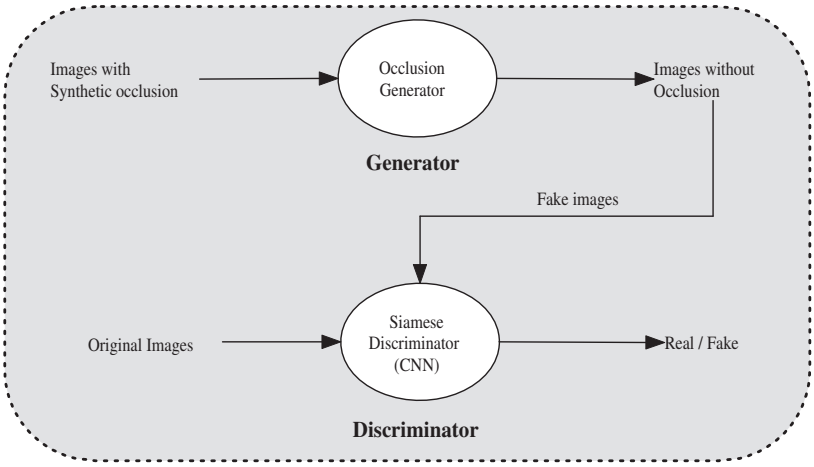

Figure 5.3: Occlusion Handling GAN (*OHGAN*) architecture

network that computes a similarity metric between the original image and the occlusion reconstructed image produced by the generator. Table 5.1 shows the layer types, filter sizes, and the number of filters in each layer of the discriminator. As seen in the table, there are five convolutional layers followed by a fully connected layer that outputs a similarity score between the original and the reconstructed image and classifies the image as real or fake by using a Sigmoid function. Additionally, in the generator architecture, there are skip connections between every convolutional layer and the corresponding de-convolutional layer. These skip connections are used to pass features from the encoder path to the decoder path in order to recover spatial information lost during down-sampling.

The training of the *OHGAN* is done using the synthetically occluded data (explained in Section 5.1). Adam optimizer [128] is used for network weight updation by fixing the momentum and learning rate to 0.5 and 0.0001, respectively. The *OHGAN* model is trained for a maximum of 20 epochs or till the loss values in two successive epochs do not undergo significant change. The L_2 loss function has been used to train this

Table 5.1: Layer specification of the *OHGAN* architecture

Network	Layers	Filter size	No. of filters
Generator	**Downsampling**		
	Conv2d_1	3×3	8
	Conv2d_2	3×3	16
	Conv2d_3	3×3	32
	Conv2d_4	3×3	64
	Upsampling		
	Deconv_4	3×3	64
	Deconv_3	3×3	32
	Deconv_2	3×3	16
	Deconv_1	3×3	3
Discriminator	Conv2d_1	5×5	20
	Conv2d_2	5×5	25
	Conv2d_3	5×5	25
	Conv2d_4	3×3	25
	Conv2d_5	3×3	25
	Fully Connected	500	-

generator. If this generator loss is denoted by L_G, then,

$$L_G = \frac{1}{CHW} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} ||G_e(I) - R_o||^2. \tag{5.2}$$

Here C, W, and H respectively represent the channel number, width, and height, $G_e(I)$ stands for the *OHGAN* generated reconstructed image and R_o represents the original image. The discriminator of *OHGAN* is a binary classifier that distinguishes between real and fake images (i.e., those generated by the generator), which is trained using binary cross-entropy loss.

As explained before, the generator of the *OHGAN* is based on a modified U-Net that consists of a multi-layer architecture of convolution and de-convolution layers with skip connections from the encoder to the decoder (refer to Table 5.1). However, the use of the skip connections may tend to retain traces of occluded patches on the generated images, thereby affecting the effectiveness of frame reconstruction. Hence, we also study the applicability of a Deep Convolutional *Autoencoder* network that does not possess any skip connections in generating these reconstructed frames. The *Autoencoder*-based occlusion reconstruction algorithm is explained next in Section 5.2.2.

5.2.2 *Autoencoder*-based Reconstruction

As with any Neural Network-based generator, an *Autoencoder* also consists of two sub-networks, one each for the encoder and the decoder. The encoder (refer to Figure 5.4) first encodes the image into a lower-dimensional latent representation, and next the decoder decodes the latent representation back to an image. In this work, we employ a Convolutional stacked *Autoencoder* due to its demonstrated effectiveness in handling image translation tasks. Similar to that of *OHGAN* discussed in Section 5.2.1, here also the input to the *Autoencoder* is the occluded image, and the output is the reconstructed unoccluded image.

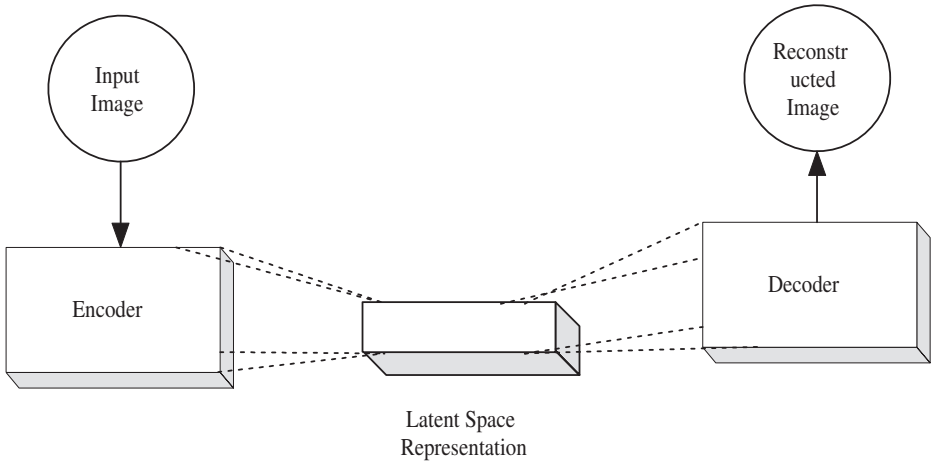

Figure 5.4: An insight view of an *Autoencoder*

The *Autoencoder* architecture used in this work consists of six layers with *ReLU* activation function except for the last layer in which Sigmoid activation has been used. The layer-wise detailed configuration of the *Autoencoder* is shown in Table 5.2. With reference to the table, the initial four layers of the *Autoencoder* are convolution layers with the same padding, kernel size of 3×3, and the number of filters in these layers are 128, 64, 64, and 32, respectively. The fifth layer is the ConvTranspose layer with the same padding, stride of 2, kernel size of 3×3, and the number of filters in this layer is 32. Finally, the last layer is a Conv2D layer with the same padding, three filters, and a

kernel size of 3×3. The model is trained using the same synthetically occluded data set that was used to train the *OHGAN*, as described in Section 5.1. Here, we use binary cross-entropy loss and Adam optimizer to update the parameters of the network. The *Autoencoder* model has been seen to achieve convergence in 1000 epochs.

Table 5.2: Layer-wise configuration of the *Autoencoder*

Network	Layer	Filter Size, Stride	No. of Filters
Encoder	Conv2d_1	3×3, 1	128
	Conv2d_2	3×3, 1	64
	Max_Pooling2d	2×2	-
	Conv2d_3	3×3, 1	64
	Conv2d_4	3×3, 1	32
Decoder	Conv2d-Transpose	3×3, 2	32
	Conv2d_1	3×3, 1	3

5.2.3 Reconstruction Results Using *OHGAN* and *Autoencoder*

The effectiveness of both the reconstruction models (i.e., *OHGAN* and *Autoencoder*) in generating occlusion-free images is evaluated using the Dice Similarity Coefficient *(DSC)* [158] (refer to Section 2.4.3 of Chapter 2). As already explained before, this metric computes the degree of spatial overlap between the two images. The *DSC* metric provides a value between *0* to *1*, with value *1* corresponding to the highest similarity and value *0* corresponding to the highest dissimilarity. First, we prepare the gallery set for training each of the occlusion reconstruction models by considering 80% training images corresponding to each subject in the *CUHK_01*, *CUHK_03*, and *Market1501* data sets (refer to Table 2.2 in Chapter 2), and corrupting these images by varying levels of synthetic occlusion. The process of construction of the synthetically occluded data has already been discussed in Section 5.1. On completion of the training phase, we obtain quite high DSC scores (>0.90) for each image used in training. The effectiveness of the trained occlusion reconstruction models on unseen data is evaluated using three other data sets, namely, *Occluded ReID*, *Partial ReID*, and *Partial iLIDS*

(refer to Section 2.4.1 of Chapter 2 for data set details). From these data sets also, we consider 80% images per class and corrupt these with varying levels of synthetic occlusion. We compute the DSC scores between the generated occlusion reconstructed images and their corresponding ground-truth occlusion-free images for each data set for the different margin values (i.e., 5, 10, 20), and report the averaged DSC score in Table 5.3 for both the *OHGAN* and *Autoencoder*. The margin value (m) indicates the maximum allowable pixel intensity difference for determining if a pixel is correctly predicted or not. Based on this, the area of overlap between the predicted and ground-truth unoccluded images is computed as per Equation 2.1 given in Chapter 2. It

Table 5.3: Dice scores obtained from the two reconstruction models at different margin values for different data sets

Data Set Name	***OHGAN***			***Autoencoder***		
	m=5	***m=10***	***m=20***	***m=5***	***m=10***	***m=20***
Occluded ReID	0.61	0.72	0.83	0.69	0.82	0.90
Partial ReID	0.69	0.76	0.88	0.75	0.89	0.94
Partial iLIDS	0.60	0.65	0.71	0.68	0.82	0.89

is observed from the table that the *Autoencoder* performs better than the *OHGAN* in terms of reconstruction capability for all the different margin values and also for the different data sets. The reason is skip connections present in the *OHGAN* tend to retain the occluded patches on the generated images as also explained in the last paragraph of Section 5.2.1, thereby reducing the *DSC* scores to a certain extent. It can be inferred from the results that if the degree of occlusion in an image is very high then the *Autoencoder*-based reconstruction will be more effective and provide more realistic results than the *OHGAN*-based reconstruction.

A qualitative comparison of the results given by the two occlusion reconstruction models has been shown in Figure 5.5. In this figure, sample synthetically occluded images are shown in the first row, whereas the *OHGAN* and *Autoencoder*-generated images are shown in the second and third rows, respectively. Visual comparison of the results

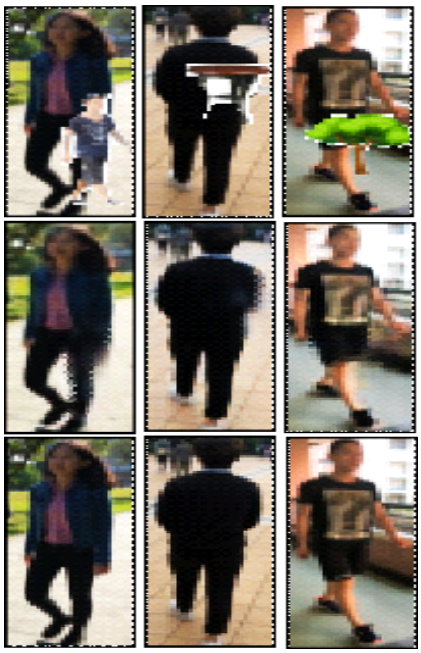

Figure 5.5: Images in the first row represent the synthetically occluded samples while the respective images in the second and third row show the generated images from the *OHGAN* and *Autoencoder*

show that the *Autoencoder*-based reconstruction results are better in terms of replacing the occluded pixels with the appropriate colors. *OHGAN* generated outputs preserve faded impressions of the occluding objects which are not present in the *Autoencoder*-generated outputs. However, the *Autoencoder*-generated outputs still contain some artifacts and noise especially at the edges of the objects in the images. To fine-tune the results further and generate more realistic occlusion reconstruction results, we propose to pass the output image from the *Autoencoder* through a Deep Convolution GAN (*DCGAN*) [159].

5.2.4 Fine-Tuning Reconstruction Results with *DCGAN*

Since in the original architecture of the *DCGAN*, the input layer of the generator accepts a random vector only, and here we are passing the *Autoencoder*-generated output through the *DCGAN* for fine-tuning, we make the required modifications to the input layer of the generator of the *DCGAN*. The layer-wise architecture of *DCGAN* is shown

in Table 5.4. As can be seen from the table, we use a combination of convolution layers,

Table 5.4: Layer-wise specification of the *DCGAN*

Network	Layer	Filter size, stride	No. of filters
Generator	Conv2d_1 to Conv2d_6	3×3, 1	32,64,64, 128,128,128
	Up_Scaling2d	2×2, 1	-
	Conv2d_7 to Conv2d_9	3×3, 1	64,64,3
Discriminator	Conv2d_1 to Conv2d_5	3×3, 2	32,64,64, 128,256
	Dense	**No. of neurons = 1**	-

up-scaling 2D layers, and batch normalization layers with $ReLU$ activation function in all but the last layer in which *tanh* activation is used. The discriminator of *DCGAN* tries to predict if the image output by the generator resembles the real (i.e., actual image) or fake (i.e., generated image). The discriminator consists of five convolutional layers with a stride value of *1* and a dropout factor of *0.2* in-between, and a final fully-connected dense layer with one neuron and Sigmoid activation function. During training the *DCGAN*, we use the *Autoencoder* generated images as input and the corresponding unoccluded ground-truth images as the desired output. It is trained with Adam optimizer in multiple epochs till convergence. During deployment, the *Autoencoder* and the *DCGAN* are connected in an end-to-end fashion, and we represent this bi-network as *Autoencoder+DCGAN*. An occluded image is input to the *Autoencoder* that does the necessary computation to generate a coarse occlusion-free image, which is then passed through the *DCGAN* to obtain the fine-tuned occlusion reconstructed image.

The reconstruction quality of the fine-tuned images is again evaluated through the Dice Similarity Coefficient *(DSC)*. Although, *Autoencoder* was seen to perform better than *OHGAN* in reconstructing the occluded frames, in this experiment we study the effect of fine-tuning both the *OHGAN* and the *Autoencoder*-generated outputs on the *DSC* score. Table 5.5 presents the *DSC* scores for the two different models, namely, *OHGAN+DCGAN* and *Autoencoder+DCGAN* for margin (m) values equal to 5, 10, and 20 using the *Partial ReID* and *Partial iLIDS* data sets. The effectiveness of *Au-*

Table 5.5: Dice similarity coefficient (DSC) values for the reconstruction models

Data set	Network	*m=5*	*m=10*	*m=20*
Partial ReID	*OHGAN*	0.68	0.73	0.79
Partial iLIDS	*+ DCGAN*	0.70	0.76	0.83
Partial ReID	*Autoencoder*	0.76	0.89	0.94
Partial iLIDS	*+ DCGAN*	0.73	0.80	0.91

toencoder over *OHGAN* in occlusion reconstruction can once again be verified from the results of this table. Moreover, on comparing Table 5.3 with Table 5.5, we observe that the DSC scores for the different margin values usually improve upon applying the fine-tuning phase. We also qualitatively study the effectiveness of the *DCGAN*-based fine-tuning stage on the reconstruction quality. For this, we observe the reconstruction results by *Autoencoder* only and also by the combination of *Autoencoder* and *DCGAN* (i.e., *Autoencoder+DCGAN*) on a set of synthetically occluded images. Figure

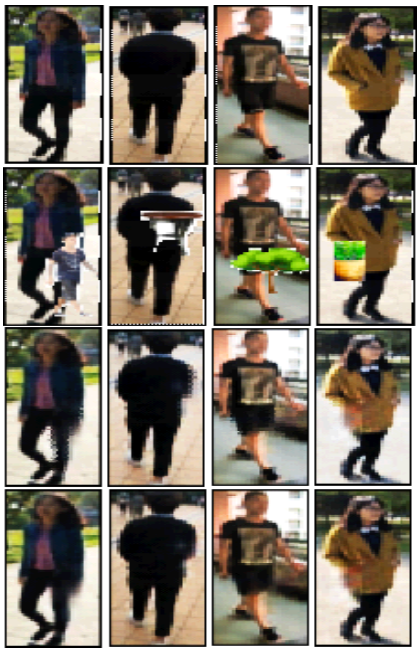

Figure 5.6: Sample ground-truth unoccluded non-sequential frames (1^{st} row), frames with synthetic occlusion (2^{nd} row), reconstruction using *Autoencoder* (3^{rd} row), and fine-tuning using *DCGAN* (4^{th} row)

5.6 presents the qualitative reconstruction results from the above two models. Here,

the first and second rows show sample image frames from the original data and the corresponding synthetically occluded frames. The third and fourth rows show the occlusion reconstructed images generated by *Autoencoder* and by *Autoencoder+DCGAN*, respectively. It can be visualized from the figure that the image frames generated by *Autoencoder+DCGAN* successfully remove the irregularities that were present on the object edges of the *Autoencoder* generated outputs, and also the generated images by *Autoencoder+DCGAN* shown in fourth row of the figure look closely similar to that of the original unoccluded images present in the data set, as shown in the first row.

5.2.5 Re-Identification Results Using Baseline Networks

The applicability of two different baseline models has been studied for person re-identification from the occlusion reconstructed data, namely, the Part-based Convolution Baseline (*PCB*), and the Siamese Convolution Box (*SCB*), introduced in Chapter 3. We compare these results with a set of non-Deep Learning-based approaches, namely, [108,147,150,160,161], and a set of Deep Learning-based approaches, namely, [3, 162, 163], and GAN-based approaches, namely, [72–74, 76] on synthetically occluded test sets generated from *CUHK_01*, *CUHK_03*, and *Market1501* data sets. The original unoccluded versions of the same 80% data corresponding to each subject used for training the occlusion reconstruction models have been also used as a gallery set for re-identification. The test set is constructed by considering the remaining 20% samples from each class and corrupting these with varying levels of synthetic occlusion. These pairs of training and test sets have been used to report the re-identification accuracy and/or map score at different ranks for the different approaches on unseen synthetically occluded data constructed from the *CUHK_01*, *CUHK_03*, and *Market1501* data sets as shown in Tables 5.6, 5.7, and 5.8, respectively. In these tables, we also present the re-identification accuracy at different ranks given by the network combinations by stacking each of *OHGAN* and *Autoencoder+DCGAN* sep-

arately with *FDGAN*, *PNGAN*, and *ResNet101*. In addition to these, we also consider a few other network combinations formed by stacking different occlusion reconstruction and re-identification models discussed in the chapter, namely, *OHGAN+PCB*, and *Autoencoder+DCGAN+PCB* for the comparative study. Results in each table are shown in four sub-blocks that correspond to the non-Deep Learning-based approaches, Deep Learning-based approaches, *GAN*-based approaches, and occlusion handling approaches, respectively. It may be noted that since the *FDGAN* and *PNGAN*-based re-identification methods do not incorporate any occlusion reconstruction mechanism, we test their efficacy in performing re-identification using the occlusion reconstructed images generated by *OHGAN/Autoencoder*.

Table 5.6: Comparative results on *CUHK_01* data set for Ranks 1, 5, and 10

Data Set	***CUHK_01***		
Methods	**Rank 1**	**Rank 5**	**Rank 10**
BoW [108]+KISSME [147]	54.8	63.0	69.0
LOMO [150]+ XQDA [150]	64.1	77.4	82.0
HistLBP [151]+XQDA [150]	56.7	70.1	81.6
WARCA [160]	56.4	66.8	74.6
Deep Re-Id [3]	31.0	44.6	56.2
SVD-Net [162]	87.2	89.5	92.0
CamStyle + re-rank [163]	86.2	88.8	93.5
MSCAN [66]	89.7	92.3	94.4
FDGAN [73]	90.4	92.7	95.6
PNGAN [74]	92.1	94.1	97.9
PT-GAN [76]	58.5	69.2	81.5
Autoencoder+DCGAN+FDGAN	90.6	93.8	97.0
Autoencoder+DCGAN+PNGAN	91.5	94.0	97.8
Autoencoder+DCGAN+ResNet101	90.6	94.4	97.8
OHGAN+FDGAN	91.2	92.8	96.0
OHGAN+PNGAN	92.4	94.8	98.0
OHGAN+ResNet101	89.0	93.4	97.8
OHGAN+PCB	93.4	93.8	96.2
Autoencoder+DCGAN+SCB	**94.0**	**96.6**	**98.2**
Autoencoder+DCGAN+PCB	93.5	94.0	95.6

From Tables 5.6, 5.7, and 5.8, it can be seen that each of the re-identification models formed by combining our proposed occlusion reconstruction and re-identification models

Table 5.7: Comparative results on *CUHK_03* data set for Ranks 1, 5, and 10

Data Set	***CUHK_03***		
Methods	**Rank 1**	**Rank 5**	**Rank 10**
BoW [108]+KISSME [147]	51.0	57.0	65.0
LOMO [150]+ XQDA [150]	66.1	75.4	82.0
HistLBP [151]+XQDA [150]	54.7	61.1	67.6
WARCA [160]	49.4	61.8	74.6
Deep Re-Id [3]	26.4	44.8	57.2
SVD-Net [162]	89.2	92.5	95.0
CamStyle + re-rank [163]	88.6	90.8	92.5
MSCAN [66]	83.7	90.3	95.4
FDGAN [73]	88.5	92.7	96.6
PNGAN [74]	90.1	93.5	97.2
PT-GAN [76]	52.5	67.2	81.5
Autoencoder+DCGAN+FDGAN	90.6	93.8	97.0
Autoencoder+DCGAN+PNGAN	91.5	94.0	**97.8**
Autoencoder+DCGAN+ResNet101	90.6	94.4	**97.8**
OHGAN+FDGAN	90.6	93.8	97.0
OHGAN+PNGAN	91.5	94.0	**97.8**
OHGAN+ResNet101	90.6	94.4	**97.8**
OHGAN+PCB	92.8	95.4	97.0
Autoencoder+DCGAN+SCB	**93.2**	**96.0**	97.4
Autoencoder+DCGAN+PCB	92.8	95.6	92.2

performs with a high accuracy/map score compared to the state-of-the-art approaches, which emphasizes the effectiveness of the proposed reconstruction models. The Rank 1 accuracy values given by the combination *Autoencoder+DCGAN+SCB* are 94%, 93.2%, and 94.4% on the synthetically occluded samples generated from *CUHK_01*, *CUHK_03*, and *Market1501* data, respectively. These accuracy values are quite satisfactory considering the fact that the input data has been corrupted with occlusion, and also these values are higher compared to that given by the other existing Deep Learning-based, GAN-based, and occlusion handling methods. The corresponding accuracy values increase to 98.2%, 97.4%, and 98.0% at Rank 10. In general, the combination *Autoencoder+DCGAN+SCB* provides the best re-identification accuracy/map score at each of the different ranks for each data set used in the study. Only for the *CUHK_03* data, it has been seen that the Rank 10 accuracy provided by each of *Au-*

Table 5.8: Comparative results on *Market1501* data set for Ranks 1, 5, and 10 along with mean average precision (*map*)

Data Set	***Market1501***			
Methods	***R-1***	***R-5***	***R-10***	***map***
BoW [108]+KISSME [147]	44.5	63.4	72.2	20.8
LOMO [150]+ XQDA [150]	32.4	44.8	60.5	17.0
HistLBP [151]+XQDA [150]	36.7	50.2	67.6	-
WARCA [160]	45.4	68.8	76.6	26.4
Deep Re-Id [3]	26.4	40.6	60.2	-
SVD-Net [162]	82.2	91.5	95.0	62.4
CamStyle + re-rank [163]	85.2	92.8	95.5	-
MSCAN [66]	86.8	-	-	66.4
FDGAN [73]	90.5	92.7	94.8	77.7
PNGAN [74]	92.0	94.5	96.8	80.9
PT-GAN [76]	40.8	-	-	30.5
Autoencoder+DCGAN+FDGAN	90.6	93.8	97.0	76.4
Autoencoder+DCGAN+PNGAN	91.5	94.0	97.8	78.0
Autoencoder+DCGAN+ResNet101	90.6	94.4	97.8	76.4
OHGAN+FDGAN	91.0	93.0	95.2	78.2
OHGAN+PNGAN	92.8	95.0	97.2	81.2
OHGAN+ResNet101	87.4	91.5	93.0	81.8
OHGAN+PCB	94.0	96.4	97.5	86.4
Autoencoder+DCGAN+SCB	**94.4**	**96.8**	**98.0**	**87.2**
Autoencoder+DCGAN+PCB	94.2	96.2	97.8	86.8

toencoder+DCGAN+PNGAN, *Autoencoder+DCGAN+ResNet101*, *OHGAN+PNGAN*, and *OHGAN+ResNet101* is 97.8%, which is slightly higher than that given by our approach, i.e., 97.4%. The map scores presented in Table 5.8 for the Market1501 data also show the effectiveness of the proposed occlusion reconstruction and re-identification models. Using the combination *Autoencoder+DCGAN+SCB*, a *map* score of 87.2 is obtained which is best among all other approaches used in the comparative study.

The results also indicate that the Siamese Convolution Box (SCB) introduced in Chapter 3, which is based on Siamese architecture, is indeed effective for image-based person re-identification, and it also performs better than PCB in each case, as can be verified from the last two rows of each table. This is mostly due to the fact that a traditional classification model like PCB requires an adequate number of examples from

each class to get trained properly. However, such extensive data is not available in the gallery set for re-identification for any of the data sets used in the study. In contrast, a sufficient quantity of training examples can be conveniently generated to train the Siamese architecture-based SCB since it only requires forming positive and negative pairs from the gallery samples. Thus, the prediction capability of SCB has also been found to be better than that of PCB. From the above results and discussions, we can conclude that the combination *Autoencoder+DCGAN+PCB* performs most accurately and robustly on various test sets compared to the other methods for image-based person re-identification in the presence of occlusion.

The same combination has also been used to test the re-identification accuracy on synthetically occluded samples present in *Partial ReID*, *Partial iLIDS*, and *IIT(BHU) Re-identification* data sets. As a re-identification gallery set we consider 80% unoccluded images corresponding to each individual from each data set, and as a test set we consider the remaining 20% unoccluded images and corrupt these with varying levels of synthetic occlusion (refer to Section 5.1). In Table 5.9, we present the corresponding Rank 1 accuracy results for each data set, and also compare these with the Rank 1 accuracy given by the two other combinations, namely, *OHGAN+PCB* and *Autoencoder+DCGAN+PCB* for each data set. Once again it can be observed from the table that *Autoencoder+DCGAN+SCB* outperforms each of the two other approaches by a significantly large margin of accuracy for each data set. For Partial ReID, the combination *Autoencoder+DCGAN+SCB* performs with 1.8% higher Rank 1 accuracy than the second-best approach, i.e., *Autoencoder+DCGAN+PCB*. The corresponding values for the Partial iLIDS and IIT (BHU) Re-Identification data sets are 1.4% and 6.6%, respectively. A closer look into the first two rows of the table also reveals that *Autoencoder+DCGAN*-based reconstruction is to some extent better than that of *OHGAN*, and this is reflected in the re-identification accuracy values as well. As in the previous experiment, the results from this table also verify the effective-

ness of *Autoencoder+DCGAN+SCB* over other models used for occlusion handling in re-identification.

Table 5.9: Comparison of Rank 1 accuracy of *OHGAN+PCB*, *Autoencoder+DCGAN+PCB*, and *Autoencoder+DCGAN+SCB* on synthetically occluded samples generated from *Partial ReID*, *Partial iLIDS*, and *IIT(BHU) Re-identification* sets

Methods	***Rank 1 accuracy (%)***		
	Partial ReID	***Partial iLIDS***	***IIT(BHU)* (*ReID Data*)**
OHGAN+PCB	75.8	67.4	89.0
Autoencoder+DCGAN+PCB	78.6	70.2	91.8
Autoencoder+DCGAN+SCB	**80.4**	**71.6**	**98.4**

In our next experiment, we make a comparative performance evaluation of the effectiveness of three different approaches in handling occluded images during re-identification. These are (i) *SCB* on occluded images without any reconstruction algorithm, (ii) the best approach determined in the previous experiments involving both reconstruction and re-identification phases, i.e., *Autoencoder+DCGAN+SCB*, and (iii) fusion of the reconstruction algorithm with part-based image analysis through SCB as discussed in Chapter 3, i.e., *Autoencoder+DCGAN+SCB (part-based)*. Rank 1 accuracy results are shown in the form of grouped bar charts in Figure 5.7 for three data sets, namely, *CUHK_01*, *CUHK_03*, and *Market1501*. In this plot, the data set names are specified in the horizontal axis and the accuracy values are plotted along the vertical axis. Each group of bars corresponds to the results obtained from the above three re-identification approaches on a given data set. From the results, it can be observed that the *SCB* without image reconstruction performs the worst among all the approaches used in the study in handling occluded images. On the other hand, by carrying out image reconstruction and re-identification using either *Autoencoder+DCGAN+SCB* or *Autoencoder+DCGAN+SCB (part-based)*, we achieve significantly higher accuracy than that provided by *SCB* alone. The results emphasize the need for employing the occlusion reconstruction technique through *Autoencoder+DCGAN* as an initial step before carrying

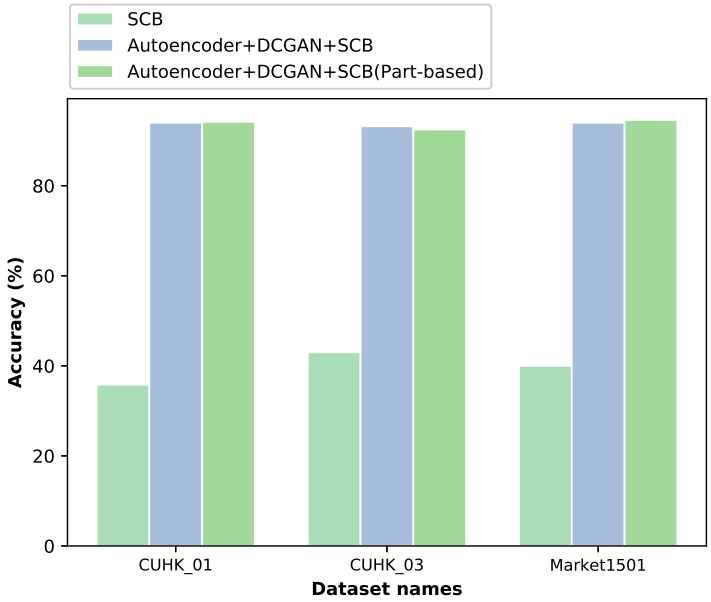

Figure 5.7: Comparison of Rank 1 accuracy of (i) *SCB*, (ii) *Autoencoder+DCGAN+SCB*, (iii) *Autoencoder+DCGAN+SCB (part-based)*

out re-identification. It can also be seen from the figure that the results obtained from *Autoencoder+DCGAN+SCB*, and *Autoencoder+DCGAN+SCB (part-based)* are almost comparable. This is mostly due to the fact that the data sets used in the experiments consist of multiple poses of the same individual from varying viewpoints, as a result of which equivalent contextual information from the images gets captured using both *SCB* on the entire image and the image part-based *SCB*. The part-based classification scheme is expected to be more effective when used in scenarios equivalent to that stated in Chapter 3, where walking images/videos of subjects are captured roughly from the same view by each camera.

5.3 Summary

In this chapter, we have discussed effective methods to handle the occlusion scenario in image-based person re-identification methods. Specifically, we have studied the applicability of two different methods, namely, *OHGAN* and *Autoencoder* for reconstructing

occlusion in images. Training of both the occlusion reconstruction models is done by constructing a data set of synthetically occluded images along with their original unoccluded versions. Qualitative and quantitative results show that among these two, the reconstruction results given by *Autoencoder* are better than that by *OHGAN*. These reconstructed results are further fine-tuned using a *DCGAN* model, and finally, the reconstructed frames corresponding to a set of occluded frames are used for re-identification. As re-identification model, we study the use of *PCB* as well as *SCB* introduced in Chapter 3. Due to the unavailability of real occluded data sets for person re-identification, as test data sets we consider the synthetically occluded versions of *CUHK_01*, *CUHK_01*, and *Market1501* data sets.

We also compare the re-identification results with several popular existing methods and found that the multi-model architecture *Autoencoder+DCGAN+SCB* performs the best for all the test data sets used in the study. Employment of the hierarchical classification scheme through color-based and SCB-based matching as discussed in Chapter 3 on the *Autoencoder+DCGAN*-reconstructed images is expected to improve the re-identification accuracy even further, which can be studied in the future. Although encouraging results are obtained using *Autoencoder+DCGAN+SCB* on occluded image data sets, it needs to be studied if spatio-temporal information-based occlusion reconstruction can help in achieving better frame reconstruction and re-identification accuracy in the case of video data sets which will be focused on in the next chapter.