Color-Sensitive Person Re-Identification

Guan'an Wang[1,3], **Yang Yang**[2], **Jian Cheng**[2,3,4], **Jinqiao Wang**[2] and *__Zengguang Hou__[1,3,4]
[1]The State Key Laboratory for Management and Control of Complex Systems,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, China
[3]University of Chinese Academy of Sciences, Beijing, China
[4]Center for Excellence in Brain Science and Intelligence Technology, Beijing, China
{wangguanan2015, zengguang.hou}@ia.ac.cn, {yang.yang, jcheng, jq.wang}@nlpr.ia.ac.cn,

Abstract

Recent deep Re-ID models mainly focus on learning high-level semantic features, while failing to explicitly explore color information which is one of the most important cues for person Re-ID model. In this paper, we propose a novel Color-Sensitive Re-ID to take full advantage of color information. On one hand, we train our model with real and fake images. By using the extra fake images, more color information can be exploited and it can avoid overfitting during training. On the other hand, we also train our model with images of the same person with different colors. By doing so, features can be forced to focus on the color difference in regions. To generate fake images with specified colors, we propose a novel Color Translation GAN (CTGAN) to learn mappings between different clothing colors and preserve identity consistency among the same person with the clothing color. Extensive evaluations on two benchmark datasets show that our approach significantly outperforms state-of-the-art Re-ID models.

1 Introduction

Person re-identification (Re-ID) aims to automatically match the underlying identities of person images from non-overlapping camera views [Gong *et al.*, 2014]. As an essential task in video surveillance of distributed multi-cameras, Re-ID is very important for individual-specific long-term behavior analysis. Due to variations of view angles, poses and illuminations in different cameras, it's very challenging to tackle this task.

In the Re-ID community, many models have been proposed, which mainly focus on three parts: hand-crafted descriptor design, metric learning and deep Re-ID models. Hand-crafted person descriptors [Satta, 2013; Zheng *et al.*, 2015a] try to design features that are robust to different view angles, poses and illuminations. Metric learning methods [Farenzena *et al.*, 2010] learn a discriminative space where the similarity of same person images is higher and that of different person images is lower. Deep Re-ID models use deep neural networks to straightly learn robust and discriminative features in an end-to-end manner. For example, [Zheng *et al.*, 2016; Hermans *et al.*, 2017] try to learn a global features by using the identity classification loss, triplet loss or quadruplet loss. [Sun *et al.*, 2018; Li *et al.*, 2018] learn part-level features based on a partition strategy or an attention mechanism. However, most of them fail to take full advantage of intelligible color information, which is one of the most important cues for person Re-ID [Kviatkovsky *et al.*, 2013; Yang *et al.*, 2014].

Figure 1: Fake images generated by our CTGAN. our CTGAN stably translates images to specified colors meanwhile maintains other content unchanged.

In this paper, we explicitly and effectively use the color information and propose a novel Color-Sensitive Re-ID model. Firstly, we learn from real and fake images by minimizing identity-based classification and triplet losses. By doing so, we can explore more color information while preserving the identity of a person. Secondly, we further take the advantage of color information by distinguishing images of the same person with different clothing colors. Specifically, we establish triplet images (x_a, x_p, x_n) from the same person. x_a and x_p own the same clothing color while x_n and x_p have different clothing colors. Thus, by minimizing the corresponding triplet loss, the learned features can focus on the color difference in regions.

*corresponding author

To generate fake images with specified colors, we also propose a novel Color Translation GAN (CTGAN), which includes a translator of generating and reconstructing images, a discriminator of distinguishing images sources and classifying clothing colors as well as a feature extractor of guaranteeing identity consistency among fake images of the same person with the same color. As shown in Figure 1, our CTGAN stably translates images to specified colors meanwhile maintains other contents unchanged.

Our main contributions can be summarized as below:

1. We propose a novel Color-Sensitive Re-ID model to explore the color information fully and explicitly. On one hand, based on the classification and triplet losses of real and fake images, more color information is exploited and the identity of a person is preserved. On the other hand, the learned features are sensitive to color changes by distinguishing images of the same person with different clothing colors.

2. We propose a novel CTGAN model to learn mappings between different clothing colors. It can stably translate person images with expected clothing colors.

3. Experiments on two publically available benchmark datasets: Market-1501 and DukeMTMC-reID demonstrate the effectiveness of our proposed Color-Sensitive Re-ID model and CTGAN. Based on them, our approach achieves new state-of-the-art results.

2 Related Works

2.1 Color Pedestrian Descriptors

In person Re-ID task, most Re-ID researchers explore color information from the person appearance and have got a series of achievements. For example, Kviatkovsky *et al.* [Kviatkovsky *et al.*, 2013] extract color features in the log-chromaticity space. Yang *et al* [Yang *et al.*, 2014] propose a novel salient color names based color descriptor to describe colors for person re-identification. Those methods prove the effectiveness of color information for person Re-ID. However, those descriptors are manually designed and hardly comparable with recent high-level semantic features of deep learning. Different from them, our model exploits color information with deep learning model.

2.2 Deep Person Re-ID Models

Deep Re-ID models automatically learn the features via deep neural networks in an end-to-end manner supervised by identity labels. Some deep Re-ID models straightly learn global feature. For example, IDE [Zheng *et al.*, 2016] learns identity-discriminative features by fine-tuning a pre-trained CNN to minimizing the identity classification loss. TriNet [Hermans *et al.*, 2017] straightly learns embeddings in a metric space by minimizing a triplet loss with batch hard. Apart from those global features Re-ID models, several part-level features Re-ID models have been proposed to learn fine-grained features via partition strategy or attention mechanism. PCB [Sun *et al.*, 2018] proposes a uniform partition on the conv-layer for learning part-level features and adaptive pooling for accurate region locations. HA-CNN [Li *et al.*, 2018] proposes a Harmonious Attention CNN to learn soft pixel attention and hard regional attention along with simultaneous optimization of the features. Different from those Re-ID models, which mainly learn abstract high-level features and fail to take full advantage of intelligible color information, our Color-Sensitive Re-ID model explicitly and effectively utilizes color information and achieves state-of-the-art performance.

Recently, some Re-ID models with GAN are also proposed. [Zheng *et al.*, 2017] show that unlabeled samples generated by GAN improve Re-ID baseline. [Zhong *et al.*, 2018] propose to learn camera-invariant descriptors with camera-style transferred images generated by StarGAN [Choi *et al.*, 2018]. Different from them, ours utilizes the idea of GAN to promote Re-ID by learning both identity-aware and color-sensitive features.

2.3 Image-to-Image Translation

Image-to-image translation is a class of vision and graphics problems where the goal is to learn mappings between(among) two(several) domains. Recently, by using a GAN loss to make the generated images and real images indistinguishable, some GAN based models [Choi *et al.*, 2018] have shown remarkable results in various computer vision tasks including image-to-image translation. Different from them, our CTGAN is specifically designed for clothing color translation, which considers an extra constraint on identity consistency of translated images from the same person and thus generating stable color translated images of a person.

3 Color-Sensitive Re-ID Model

3.1 Learning from Real Images (Baseline)

The classification loss and the triplet loss have been proven to be efficient in learning identity-discriminative features for Re-ID. Inspired by [Zheng *et al.*, 2016; Hermans *et al.*, 2017], we use a CNN model to learn the feature map M of an image and then pool it to a feature vector V. Given real images X^r, we optimize their feature vectors with the classification loss $\mathcal{L}_{cls}^{real}$ of a classifier C and the triplet loss $\mathcal{L}_{trl}^{real}$ of an embedder E.

$$\begin{aligned}\mathcal{L}_{basel.} &= \mathcal{L}_{cls}^{real} + \mathcal{L}_{trl}^{real} \\ &= E_{x\in X^r}[-log\ p(x)] + \mathcal{L}_{trl}(X^r, X^r, X^r)\end{aligned} \tag{1}$$

where $p(\cdot)$ is the predicted probability of the input belonging to the ground-truth. $\mathcal{L}_{trl}(\cdot,\cdot,\cdot)$ is defined in Eq.(2), where x_a and x_p are a positive pair belonging to the same person, x_a and x_n are a negative pair belonging to different persons, D_{x_1,x_2} is the cosine distance between x_1 and x_2 in embedding space of the embedder E, and m is a margin parameter, $[x]_+ = max(0, x)$.

$$\begin{aligned}&\mathcal{L}_{trl}(X_1, X_2, X_3) = [m - D_{x_a,x_p} + D_{x_a,x_n}]_+ \\ &s.t.\ \ x_a \in X_1, x_p \in X_2, x_n \in X_3\end{aligned} \tag{2}$$

3.2 Extension of Baseline to Fake Images

In Eq. (1), features learned from real images do not explicitly take full advantage of color information. To solve it, we generate more fake images with different colors by the CTGAN

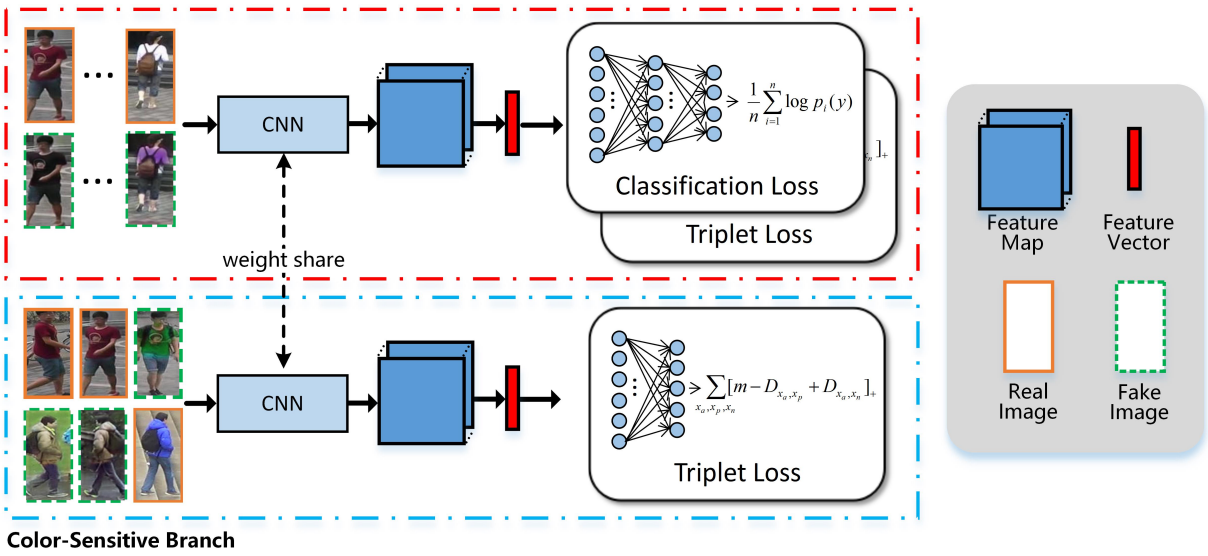

Figure 2: Architecture of our proposed Color-Sensitive Re-ID model. More details can be found in text.

(please see Section 4) and add them to train our baseline Re-ID model under a triplet loss. Then, the newly added triplets are constructed as follows: positive pairs are from real (or fake) images of the same person while the negative sample correspondingly comes from the fake (or real) image of a different person with the anchor. Considering that fake images may have noise, we exclude the triplet using all fake images. Then, our extension of baseline to fake images can be formulated as:

$$\mathcal{L}_{real+fake} = \mathcal{L}_{cls}^{real} + \mathcal{L}_{trl}^{real+fake} \tag{3}$$

Compared with Eq. (1), Eq. (3) further explores more color information among existing persons. It brings the following advantages: (1) we have extra positive pairs with different clothing colors, thus making the model be adaptive with different colors; (2) we have extra negative pairs and may make it easier to distinguish different person with similar clothing color; (3) the augmented data is generated without extra annotations and it can avoid over-fitting during training.

3.3 Learning from the Same Person with Different Colors

Although we can learn more color information with $\mathcal{L}_{real+fake}$ in Eq.(3) than $L_{basel.}$ in Eq.(1), color information is not fully exploited because it does not consider the case with different clothing colors of the same person. To that end, we propose a novel color-sensitive branch where triplets are constructed with real and fake images of the same person. Specifically, for a person, we choose images with the same clothing color as positive pairs and treat those with different clothing colors as the negative pairs. Then, for each person i, the corresponding loss can be formulated as below:

$$\mathcal{L}_{color}^{i} = \mathcal{L}_{trl}(X_i^r, X_i^r, X_i^f) + \mathcal{L}_{trl}(X_i^f, X_i^f, X_i^r) \tag{4}$$

where X_i^r is all real images of the ith person and X_i^f denotes all fake images of the ith person. The final loss of our color-sensitive branch can be formulated as below:

$$\mathcal{L}_{color} = E_i[\mathcal{L}_{color}^{i}] \tag{5}$$

Based on Eq. (5), we can take full advantage of color information and make the learned features focus on clothing color changes.

3.4 Overall Objective Function

Finally, the overall objective function of our Color-Sensitive Re-ID model is formulated in Eq.(6). Our whole Color-Sensitive Re-ID model can be optimized by scholastic gradient descent in an end-to-end way.

$$\mathcal{L}_{all} = \mathcal{L}_{cls}^{real} + \mathcal{L}_{trl}^{real+fake} + \mathcal{L}_{color} \tag{6}$$

3.5 Discussion

Although both $\mathcal{L}_{real+fake}$ in Eq.(3) and $\mathcal{L}_{color}$ in Eq.(5) use fake images, they are total different. The former constructs triplets with different persons, which contributes to exploring more color information. The latter constructs a triplet using images of the same person, which makes the learned features focus on clothing color changes. The experimental results in Section 5.3 show the effectiveness of both contributions.

4 Color Translation GAN (CTGAN)

To generate fake images with specified clothing colors, we propose a novel Color Translation Generative Adversarial Network (CTGAN). As shown in Figure 3, the CTGAN consists of three components, *i.e.* a translator T, a discriminator D and a feature extractor F. The module T is to translate a x to a fake image x' with color c', *i.e.* $x' = T(x, c')$. The module D is to distinguish images source (real or fake images) $D_{src}(\cdot)$ and classify images color $D_{cls}(\cdot)$. The module F extracts features to compute the similarity between any two images.

Specifically, the objective function of our CTGAN includes an adversarial loss $\mathcal{L}_{adv}$, a color classification loss $\mathcal{L}_{cls}$, a reconstruction loss $\mathcal{L}_{rec}$, and an identity consistency loss $\mathcal{L}_{icl}$. $\mathcal{L}_{adv}$ is to make fake images as realistic as possible by making fake images indistinguishable from real ones. $\mathcal{L}_{cls}$ is to translate images to a specified color. This is implemented by making $D_{cls}(\cdot)$ classify a real image x to its ground truth color c, and T translates an image x to x' that can be classified as a specified color c'. $\mathcal{L}_{rec}$ guarantees that the fake image x' preserves the contents of a person by making the x' reconstruct corresponding real ones x. Thus, $\mathcal{L}_{adv}$, $\mathcal{L}_{cls}$ and $\mathcal{L}_{rec}$ can be formulated in Eq.(7), Eq.(8) and Eq.(9).

$$\begin{aligned}\min_{T}\max_{D_{src}} \mathcal{L}_{adv} &= E_x[log\ D_{src}(x)] \\ &+ E_{x,c'}[log(1 - D_{src}(T(x, c')))]\end{aligned} \tag{7}$$

$$\begin{aligned}\min_{D_{cls}} \mathcal{L}^D_{cls} &= E_{(x,c)}[-log\ D_{cls}(c|x)] \\ \min_{T} \mathcal{L}^T_{cls} &= E_{x,c'}[-log\ D_{cls}(c'|(T(x, c')))]\end{aligned} \tag{8}$$

$$\min_{T} \mathcal{L}_{rec} = E_{(x,c),c'}[||x - T(T(x, c'), c)||_1] \tag{9}$$

Although the above-mentioned three losses learn multi-domain mappings between different clothing colors, it cannot generate stable clothing color translated images of a person. This will be shown in Section 5.4. To solve the problem, we propose a novel identity-consistency loss in Eq.(10).

$$\max_{T} \mathcal{L}_{icl} = E_{(x_1,x_2)}[D_{x_1,x_2}] \tag{10}$$

where x_1 and x_2 are the fake images of the same person with the same clothing color, D_{x_1,x_2} means the cosine distance between two images in the feature space learned by the module F. Note that the module F is a pre-trained feature learning part for person Re-ID and is fixed when training our CTGAN. With it, we can guarantee that the fake images of a person with the same clothing color stay close in the feature space and thus translating person images with specified colors steadily.

The overall objective function of our proposed CTGAN model can be written in Eq.(11), where $\lambda_{cls}, \lambda_{rec}, \lambda_{icl}$ are trade-off weights of $\mathcal{L}_{cls}$, $\mathcal{L}_{rec}$ and $\mathcal{L}_{icl}$, respectively. We set $\lambda_{cls} = 1, \lambda_{rec} = 10$ as in [Choi *et al.*, 2018] and $\lambda_{icl} = 1$ empirically. Our CTGAN can be learned by alternatively optimizing $\mathcal{L}_D$ and $\mathcal{L}_T$.

$$\begin{aligned}\min_{D} \mathcal{L}_D &= -\mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}^D_{cls} \\ \min_{T} \mathcal{L}_T &= \mathcal{L}_{adv} + \lambda_{cls}\mathcal{L}^T_{cls} + \lambda_{rec}\mathcal{L}_{rec} - \lambda_{icl}\mathcal{L}_{icl}\end{aligned} \tag{11}$$

5 Experiment

5.1 Datasets

Market-1501 [Zheng *et al.*, 2015b] contains 32,669 annotated images of 1,501 identities from 6 cameras. Each identity is annotated 8 colors of upper-body clothing and 9 colors of lower-body clothing. DukeMTMC-reID [Ristani *et al.*, 2016] includes 16,522 training images of 702 identities. Each identity is annotated 8 colors of upper-body clothing and 7 colors of lower-body clothing. The color attributes of both datasets are annotated by [Lin *et al.*, 2017].

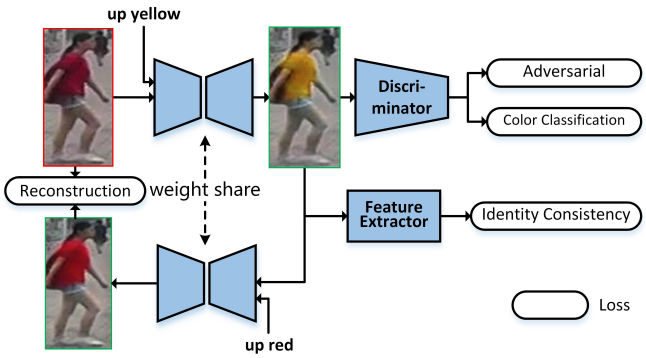

Figure 3: Architecture of our proposed Color Translation GAN (CTGAN). Our CTGAN learns mappings between any two different colors, which translates images with the specified color.

5.2 Implementation Details

For the CNN part, we adopt ResNet-50 [He *et al.*, 2016] pre-trained on ImageNet [Deng *et al.*, 2009] as our CNN backbone, and use the Pool5 layer as our feature map M. For the classification loss, a classifier includes a 256-dim FC layer as the middle layer followed by batch normalization and dropout, and an FC layer with identity number logits as output layer. For the triplet loss, an embedder is an FC layer which maps a part-level feature vector to 256-dimensions embeddings.

Recent works [Li *et al.*, 2018; Sun *et al.*, 2018] show that part-level features own more fine-grained information, which contributes to accurate Re-ID. As in [Sun *et al.*, 2018], we uniformly partition the feature map M into 6 horizontal stripes, pooling them into 6 part-level vectors $\{V_i\}^6_{i=1}$, and optimize them with non-shared classifiers and embedders. Considering that some parts of negative pairs may be unchanged after CTGAN, from which we can hardly learn color information, we abandon those parts of negative pairs with cosine distance larger than a constant (we set 0.6 throughout the paper). During the test stage, we use the concatenation of normalized feature vectors $\{V_i\}^6_{i=1}$ as person features and the cosine distance as similarity.

We implement our approach with Pytorch. The training images are augmented with horizontal flip, random cropping, random erasing [Zhong *et al.*, 2017] and normalization. The batch size of real images is set to 192 (24 persons and 8 images for a person), and that of fake images is set to 112 (4 persons, randomly select 7 fake images for a person). We initialize the learning rates of the CNN part at 0.05 and the other parts (classifiers and embedders) at 0.5. The learning rates are decayed by 0.1 every $4,000$ iterations, and the model is trained for $15,000$ iterations in total. We set margin $m = 0.1$ empirically.

5.3 Ablation Study

To evaluate each component of our color-sensitive Re-ID model, we conduct two variants trained under different settings. Firstly, we train the model under Eq.(1) as a baseline. Secondly, we evaluate the effectiveness of the learning from

Methods	Market-1501				DukeMTMC-reID			
	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10	mAP
$\mathcal{L}_{basel.}$	93.1	97.6	98.4	80.5	82.3	91.2	93.1	68.5
$\mathcal{L}_{real+fake}$	94.5	98.0	98.6	82.6	83.9	91.9	93.9	69.9
Ours($\mathcal{L}_{all}$)	97.0	98.3	98.7	87.1	86.7	92.7	94.6	73.9

Table 1: Comparison with different variants of our Color-Sensitive Re-ID model on Market-1501 and DukeMTMC-reID.

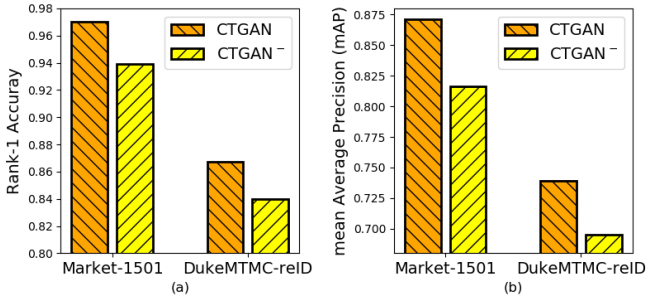

Figure 4: Experimental results of Color-Sensitive Re-ID model when using CTGAN and CTGAN$^-$.

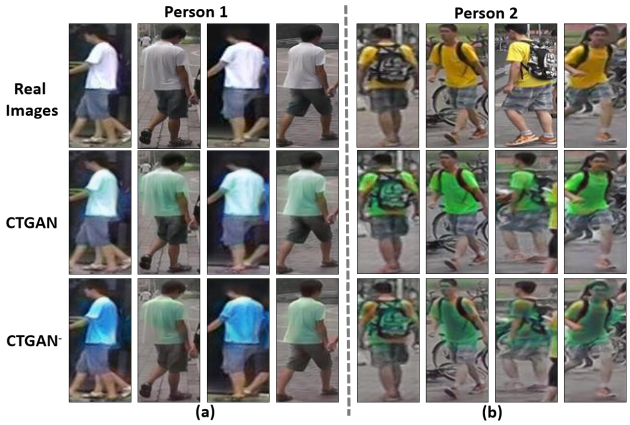

Figure 5: Comparison between fake images generated by CTGAN and CTGAN$^-$. CTGAN generates stable images while CTGAN$^-$ fails to. (a) Colors of fake images of a person generated by CTGAN are relatively stable. (b) CTGAN accurately finds clothing regions and change the color, but CTGAN$^-$ fails to.

both real and fake images under Eq.(3). Thirdly, by comparing with our whole model under Eq.(6) with the first two variants, we can verify the effectiveness of learning by distinguishing the same person with different colors. We mark them as $\mathcal{L}_{basel.}$, $\mathcal{L}_{real+fake}$ and *Ours*($\mathcal{L}_{all}$), respectively.

As shown in Table 1, $\mathcal{L}_{basel.}$ achieves 80.5% and 68.5% mAP scores and 93.1% and 82.3% Rank-1 scores on Market-1501 and DukeMTMC-reID, respectively. This demonstrates that the identity-based classification and triplet losses of real images are able to learn identity-discriminative features for person Re-ID. Secondly, $\mathcal{L}_{real+fake}$ achieves 82.6% and 69.9% mAP scores and 94.5% and 83.9% Rank-1 scores on Market-1501 and DukeMTMC-reID, respectively. Compared with $\mathcal{L}_{basel.}$, the fake images improve the results by 2.1% and 1.4% mAP scores and 1.4% and 1.6% Rank-1 scores on Market-1501 and DukeMTMC-reID, respectively. This demonstrates that the fake images generated by our CTGAN are meaningful and contribute to learn color-sensitive features. Finally, our whole Color-Sensitive Re-ID model further improves $\mathcal{L}_{real+fake}$ by 4.5% and 4.0% mAP scores and 2.5% and 2.8% Rank-1 scores on Market-1501 and DukeMTMC-reID, respectively. This significant improvement indicates that color information is critical to improve the discriminative ability and distinguishing the same person with different color is an effective way to make the features more discriminative. In summary, *Ours*($\mathcal{L}_{all}$) significantly improves $\mathcal{L}_{basel.}$ by 6.6% and 5.4% mAP scores and 3.9% and 4.4% Rank-1 scores on Market-1501 and DukeMTMC-reID, respectively. This verifies the effectiveness of our proposed approach by using extra color information.

5.4 CTGAN Evaluation

Fig.1 displays the fake images generated by our Color Translation GAN (CTGAN). As we can see that, our CTGAN stably translates real images of a person by using expected clothing colors and keeps other contents unchanged.

We also compare fake images generated by our CTGAN with and without identity-consistency loss $\mathcal{L}_{icl}$ in Eq.(10). We mark the CTGAN without $\mathcal{L}_{icl}$ as CTGAN$^-$. As is seen, our CTGAN generates stable images while the CTGAN$^-$ can not. For example, Fig.5(a) shows that the color of fake images of a person generated by CTGAN is relatively stable, but those generated by CTGAN$^-$ are biased. Fig.5(b) demonstrates that our CTGAN$^-$ accurately finds clothing regions and changes the color, but the CTGAN$^-$ fails to.

The unstable fake images may import noise to the following Color-Sensitive Re-ID model. We compare the results of our Color-Sensitive Re-ID model with fake images generated by CTGAN and CTGAN$^-$ in Fig.4. We can see that, when using CTGAN$^-$, the performance drops a lot. Specifically, there are declines of 3.1% and 2.7% Rank-1 scores and 5.5% and 4.4% mAP scores on Market-1501 and DukeMTMC-reID, respectively. This phenomenon demonstrates the effectiveness of our CTGAN from a point of quantitative view.

5.5 Effect of Different Clothing Regions

To analyze the effect of different clothing regions, we also report the experimental results on Market-1501 when using fake images generated by different upper or lower clothing colors. We represent them as *Ours*($\mathcal{L}_{lower}$) and *Ours*($\mathcal{L}_{upper}$). As we can see in Table 2, when the fake images with different upper or lower clothing colors are used, better results can be achieved than baseline. In addition, the

Methods	Market-1501			
	Rank-1	Rank-5	Rank-10	mAP
$\mathcal{L}_{basel.}$	93.1	97.6	98.4	80.5
Ours($\mathcal{L}_{lower}$)	95.5	98.0	98.6	84.9
Ours($\mathcal{L}_{upper}$)	96.2	98.2	98.6	86.1
Ours($\mathcal{L}_{all}$)	97.0	98.3	98.7	87.1

Table 2: Performance on Market-1501 when generating fake images with upper or lower clothing colors.

Backbones	$\mathcal{L}_{basel.}$		*Ours*($\mathcal{L}_{all}$)	
	Rank-1	mAP	Rank-1	mAP
AlexNet	79.5	60.2	84.6	68.8
VGG16	87.1	73.1	93.2	79.9
ResNet-50	93.1	80.5	97.0	87.1

Table 3: Peformance on Market-1501 when using different backbones. Our approach works well with different CNN backbones.

upper part seems more important than lower one and we can obtain the best results by using both.

5.6 Performance with Different CNN Backbones

As we all know, the architecture of CNN has a large effect on various machine learning tasks [Krizhevsky *et al.*, 2012; Simonyan and Zisserman, 2015; He *et al.*, 2016]. To verify the robustness of our approach with different CNN architectures, we evaluate our approach with three different CNN backbones, including AlexNet [Krizhevsky *et al.*, 2012], VGG16 [Simonyan and Zisserman, 2015] and ResNet-50 [He *et al.*, 2016]. The experimental results on Market-1501 are shown in Table 3. As we can see, *Ours*($\mathcal{L}_{all}$) improves $\mathcal{L}_{basel.}$ by 8.6%, 6.8% and 6.6% mAP scores on Market-1501 with AlexNet, VGG16 and ResNet-50 respectively. This verifies the robustness of our proposed approach to different CNN backbones.

5.7 Comparison with State-of-the-art

We compare our Color-Sensitive Re-ID model with state-of-the-art, which can be grouped into three groups, *i.e.* hand-crafted descriptors, deep Re-ID models of global features, deep Re-ID models of local-level features. All deep Re-ID models are based on ResNet-50.

The experimental results on Market-1501 and DukeMTMC-reID are shown in Table 4 and Table 5. Firstly, the hand-crafted descriptors contribute to re-identification but achieve an unsatisfactory performance. This is because hand-crafted descriptors are usually designed manually without identity labels. Secondly, deep Re-ID models with global features significantly outperform the hand-crafted descriptors, which demonstrates the effectiveness of deep learning. Thirdly, deep Re-ID models with part-level features further improve those of global features, which demonstrates that part-level features acquire fine-grained information and can be in favor of accurate re-identification. Finally, Ours outperforms all those methods achieving new state-of-the-art. The experimental results verify the superiority of our Color-Sensitive Re-ID model over existing methods.

	Methods	Rank-1	Rank-5	mAP
1	BoW+kissme [Satta, 2013]	44.4	63.9	20.8
	KLFDA [Zheng *et al.*, 2015a]	46.5	71.1	-
2	IDE [Zheng *et al.*, 2016]	85.3	-	68.5
	SVDNet [Sun *et al.*, 2017]	82.3	92.3	62.1
	Zheng [Zheng *et al.*, 2017]	83.9	-	66.1
	TriNet [Hermans *et al.*, 2017]	84.9	94.2	69.1
	DML [Zhang *et al.*, 2018]	87.7	-	68.8
	CamStyle [Zhong *et al.*, 2018]	88.1	-	68.7
3	MultiScale [Chen *et al.*, 2017]	88.9	-	79.1
	GLAD [Wei *et al.*, 2017]	89.9	-	73.9
	HA-CNN [Li *et al.*, 2018]	91.2	-	75.7
	PCB [Sun *et al.*, 2018]	92.3	97.2	77.4
	PCB+RPP [Sun *et al.*, 2018]	93.8	97.5	81.6
	Ours($\mathcal{L}_{all}$)	**97.0**	**98.5**	**87.1**

Table 4: Comparison with state-of-the-art Re-ID models on Market-1501. 1: hand-crafted descriptors. 2: deep Re-ID models with global features. 3: deep Re-ID models with part-level features.

	Methods	Rank-1	mAP
1	BoW+kissme [Satta, 2013]	25.1	12.2
	LOMO+XQDA [Liao *et al.*, 2015]	30.8	17.0
2	IDE [Zheng *et al.*, 2016]	65.2	45.0
	Zheng *et al.* [Zheng *et al.*, 2017]	67.7	47.2
	SVDNet [Sun *et al.*, 2017]	76.7	56.8
	CamStyle [Zhong *et al.*, 2018]	75.3	53.5
3	MultiScale [Chen *et al.*, 2017]	79.2	60.6
	HA-CNN [Li *et al.*, 2018]	80.5	63.8
	PCB [Sun *et al.*, 2018]	81.8	66.1
	PCB+RPP [Sun *et al.*, 2018]	83.3	69.2
	Ours($\mathcal{L}_{all}$)	**86.7**	**73.9**

Table 5: Comparison with state-of-the-art Re-ID models on DukeMTMC-reID. 1: hand-crafted descriptors. 2: deep Re-ID models with global features. 3: deep Re-ID models with local features.

6 Conclusion

In this paper, we propose a novel approach to exploit color-sensitive features for person Re-ID. Firstly, we propose a novel Color-Sensitive Re-ID model, which learns color information by learning from real and fake images, and distinguishing images of the same person with different clothing colors. Secondly, we design a novel Color Translation GAN (CTGAN) to learn mappings between different colors, where an identity consistency is proposed to improve the stability of generating fake images of a person. Finally, experimental results on two benchmark datasets show the effectiveness.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 61720106012, 61533016 and 61806203, the Strategic Priority Research Program of Chinese Academy of Science under Grant XDBS01000000, the Beijing Natural Science Foundation under Grant L172050.

References

[Chen *et al.*, 2017] Yanbei Chen, Xiatian Zhu, and Shaogang Gong. Person re-identification by deep learning multi-scale representations. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2590–2600, 2017.

[Choi *et al.*, 2018] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. *computer vision and pattern recognition*, pages 8789–8797, 2018.

[Deng *et al.*, 2009] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.

[Farenzena *et al.*, 2010] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino, and Marco Cristani. Person re-identification by symmetry-driven accumulation of local features. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2360–2367, 2010.

[Gong *et al.*, 2014] Shaogang Gong, Marco Cristani, Shuicheng Yan, and Chen Change Loy. *Person re-identification*. Springer, 2014.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[Hermans *et al.*, 2017] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*, 2017.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[Kviatkovsky *et al.*, 2013] Igor Kviatkovsky, Amit Adam, and Ehud Rivlin. Color invariants for person reidentification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7):1622–1634, 2013.

[Li *et al.*, 2018] Wei Li, Xiatian Zhu, and Shaogang Gong. Harmonious attention network for person re-identification. In *CVPR*, volume 1, page 2, 2018.

[Liao *et al.*, 2015] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z Li. Person re-identification by local maximal occurrence representation and metric learning. In *IEEE conference on computer vision and pattern recognition*, pages 2197–2206, 2015.

[Lin *et al.*, 2017] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, and Yi Yang. Improving person re-identification by attribute and identity learning. *arXiv preprint arXiv:1703.07220*, 2017.

[Ristani *et al.*, 2016] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision workshop on Benchmarking Multi-Target Tracking*, 2016.

[Satta, 2013] Riccardo Satta. Appearance descriptors for person re-identification: a comprehensive review. *arXiv preprint arXiv:1307.5748*, 2013.

[Simonyan and Zisserman, 2015] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *international conference on learning representations*, 2015.

[Sun *et al.*, 2017] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3820–3828, 2017.

[Sun *et al.*, 2018] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *ECCV*, 2018.

[Wei *et al.*, 2017] Longhui Wei, Shiliang Zhang, Hantao Yao, Wen Gao, and Qi Tian. Glad: Global-local-alignment descriptor for pedestrian retrieval. In *ACM on Multimedia Conference*, pages 420–428, 2017.

[Yang *et al.*, 2014] Yang Yang, Jimei Yang, Junjie Yan, Shengcai Liao, Dong Yi, and Stan Z. Li. Salient color names for person re-identification. In *European Conference on Computer Vision*, pages 536–551, 2014.

[Zhang *et al.*, 2018] Ying Zhang, Tao Xiang, Timothy M. Hospedales, and Huchuan Lu. Deep mutual learning. *computer vision and pattern recognition*, pages 4320–4328, 2018.

[Zheng *et al.*, 2015a] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1116–1124, 2015.

[Zheng *et al.*, 2015b] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Computer Vision, IEEE International Conference on*, 2015.

[Zheng *et al.*, 2016] Liang Zheng, Yi Yang, and Alexander G Hauptmann. Person re-identification: Past, present and future. *arXiv preprint arXiv:1610.02984*, 2016.

[Zheng *et al.*, 2017] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3774–3782, 2017.

[Zhong *et al.*, 2017] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*, 2017.

[Zhong *et al.*, 2018] Zhun Zhong, Liang Zheng, Zhedong Zheng, Shaozi Li, and Yi Yang. Camera style adaptation for person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5157–5166, 2018.