Deep Reinforcement Learning Attention Selection for Person Re-Identification

Xu Lan[1]
x.lan@qmul.ac.uk

Hanxiao Wang[2]
hxw@bu.edu

Shaogang Gong[1]
s.gong@qmul.ac.uk

Xiatian Zhu[1]
xiatian.zhu@qmul.ac.uk

[1] Computer Vision Group,
School of Electronic Engineering and
Computer Science,
Queen Mary University of London,
London E1 4NS, UK

[2] Electrical and Computer Engineering
Department,
Boston University,
Boston MA 02215, USA

Abstract

Existing person re-identification (re-id) methods assume the provision of accurately cropped person bounding boxes with minimum background noise, mostly by *manually cropping*. This is significantly breached in practice when person bounding boxes must be *detected automatically* given a very large number of images and/or videos processed. Compared to carefully cropped manually, auto-detected bounding boxes are far less accurate with random amount of background clutter which can degrade notably person re-id matching accuracy. In this work, we develop a joint learning deep model that optimises person re-id attention selection within any auto-detected person bounding boxes by *reinforcement learning* of background clutter minimisation subject to re-id label pairwise constraints. Specifically, we formulate a novel unified re-id architecture called **I**dentity **D**iscriminativ**E A**ttention reinforcement **L**earning (IDEAL) to accurately select re-id attention in auto-detected bounding boxes for optimising re-id performance. Our model can improve re-id accuracy comparable to that from exhaustive human manual cropping of bounding boxes with additional advantages from identity discriminative attention selection that specially benefits re-id tasks beyond human knowledge. Extensive comparative evaluations demonstrate the re-id advantages of the proposed IDEAL model over a wide range of state-of-the-art re-id methods on two auto-detected re-id benchmarks CUHK03 and Market-1501.

1 Introduction

Person re-identification (re-id) aims at searching people across non-overlapping camera views distributed at different locations by matching person bounding box images [14]. In real-world re-id scenarios, *automatic person detection* [12] is *essential* for re-id to scale up to large size data, e.g. more recent re-id benchmarks CUHK03 [23] and Market-1501 [69]. Most existing re-id test datasets (Table 1) are *manually cropped*, as in VIPeR [15] and
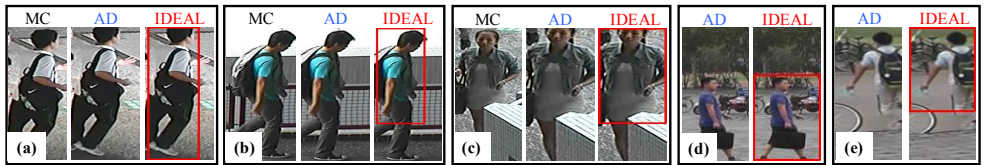
Figure 1: Comparisons of person bounding boxes by manually cropping (MC), automatically detecting (AD), and identity discriminative attention reinforcement learning (IDEAL). Often AD contains more background clutter (a,d,e). Both AD and MC may suffer from occlusion (c), or a lack of identity discriminative attention selection (b).

iLIDS [41], thus they do not fully address the re-id challenge in practice. However, auto-detected bounding boxes are not optimised for re-id tasks due to potentially more background clutter, occlusion, missing body part, and inaccurate bounding box alignment (Fig. 1). This is evident from that the rank-1 re-id rate on CUHK03 drops significantly from 61.6% on manually-cropped to 53.4% on auto-detected bounding boxes by state-of-the-art hand-crafted models [54], that is, a 8.2% rank-1 drop; and from 75.3% on manually-cropped [62] to 68.1% on auto-detected [50] by state-of-the-art deep learning models, that is, a 7.2% rank-1 drop. Moreover, currently reported "auto-detected" re-id performances on both CUHK03 and Market-1501 have further benefited from artifical *human-in-the-loop cleaning process*, which discarded "bad" detections with < 50% IOU (intersection over union) overlap with corresponding manually cropped bounding boxes. Poorer detection bounding boxes are considered as "distractors" in Market-1501 and not given re-id labelled data for model learning. In this context, there is a need for *attention selection* within auto-detected bounding boxes as an integral part of learning to optimise person re-id accuracy in a fully automated process.

Table 1: Six benchmarking person re-identification datasets with/without auto-detection introduced in the past decade. MC: Manual Cropping; AD: Automatic Detection.

Dataset	VIPeR [15]	GRID [40]	iLIDS [41]	CAVIAR4ReID [8]	CUHK03 [23]	Market-1501 [69]
Year	2007	2009	2010	2011	2014	2015
Annotation	MC	MC	MC	MC	MC+AD	AD
Identities	632	250	119	72	1,360	1,501
Images	1,264	1,275	476	1,221	28,192	32,668

There is very little attempt in the literature for solving this problem of attention selection within auto-detected bounding boxes for optimising person re-id, except a related recent study on joint learning of person detection and re-id [53]. Our approach however differs from that by operating on any third party detectors *independently* so to benefit continuously from a wide range of detectors being rapidly developed by the wider community. Other related possible strategies include local patch calibration for mitigating misalignment in pairwise image matching [44, 50, 57, 72] and local saliency learning for region soft-selective matching [27, 53, 67, 68]. These methods have shown to reduce the effects from viewpoint and human pose change on re-id accuracy. However, *all* of them assume that person bounding boxes are reasonably accurate.

In this work, we consider the problem of optimising attention selection within any auto-detected person bounding boxes for maximising re-id tasks. The **contributions** of this study are: **(1)** We formulate a novel **I**dentity **D**iscriminativ**E** **A**ttention reinforcement **L**earning (IDEAL) model for attention selection *post-detection* given re-id discriminative constraints. Specifically, IDEAL is designed to locate automatically identity-sensitive attention regions within auto-detected bounding boxes by optimising recursively attending actions using *rein-*

forcement learning subject to a reward function on satisfying re-id pairwise label constraints (Fig. 2). In contrast to existing saliency selection methods, this global attention selection approach is more scalable in practice. This is because that most saliency models are local-patch based and assume good inter-image alignment, or it requires complex manipulation of local patch correspondence *independently*, difficult to scale. The IDEAL attention model is directly estimated under a discriminative re-id matching criterion to jointly maximise a reinforcement agent model by learning *reward* it experiences. Moreover, the IDEAL attention selection strategy has the flexibility to be readily integrated with different deep learning features and detectors therefore can benefit directly from models rapidly developed elsewhere. **(2)** We introduce a simple yet powerful deep re-id model based on the Inception-V3 architecture [48]. This model is learned directly by the identity classification loss rather than the more common pairwise based verification [2, 23] or triplet loss function [11]. This loss selection not only significantly simplifies training data batch construction (e.g. random sampling with no notorious tricks required [22]), but also makes our model more scalable in practice given a large size training population or imbalanced training data from different camera views. We conducted extensive experiments on two large auto-detected datasets CUHK03 [23] and Market-1501 [69] to demonstrate the advantages of the proposed IDEAL model over a wide range (24) of contemporary and state-of-the-art person re-id methods.

2   Related Work

Most existing re-id methods [7, 11, 21, 23, 25, 32, 36, 39, 55, 56, 57, 58, 71] focus on supervised learning of person identity-discriminative information. Representative learning algorithms include ranking by pairwise constraints [31, 38, 57, 59], discriminative sub-space/distance metric learning [7, 21, 26, 39, 40, 64, 65, 71], and deep learning [2, 11, 11, 23, 24, 45, 52, 62]. They typically require a large quantity of person bounding boxes and inter-camera pairwise identity labels, which is prohibitively expensive to collect manually.

Automatic Detection in Re-ID   Recent works [23, 69, 70, 70] have started to use automatic person detection for re-id benchmark training and test. Auto-detected person bounding boxes contain more noisy background and occlusions with misaligned person cropping (Fig. 1), impeding discriminative re-id model learning. A joint learning of person detection and re-id was also investigated [53]. However, the problem of *post-detection* attention selection for re-id studied in this work has not been addressed in the literature. Attention selection can benefit independently from detectors rapidly developed by the wider community.

Saliency and Attention Selection in Re-ID   Most related re-id techniques are localised patch matching [44, 50, 72] and saliency detection [27, 53, 67, 68]. They are inherently unsuitable by design to cope with poorly detected person images, due to their stringent requirement of tight bounding boxes around the whole person. In contrast, the proposed IDEAL model is designed precisely to overcome inaccurate bounding boxes therefore can potentially benefit all these existing methods.

Reinforcement Learning in Computer Vision   Reinforcement Learning (RL) [57] is a problem faced by an agent that learns its optimal behaviour by trial-and-error interactions with a dynamic environment [18]. The promise of RL is offering a way of guiding the agent learning by reward and punishment without the need for specifying how the target tasks to be realised. Recently, RL has been successfully applied to a few vision tasks such as object localisation [3, 5, 17, 35], image captioning [28, 43], active object recognition [33]. To our best knowledge, this is the first attempt to exploit reinforcement learning for person re-id. Compared to the most related fully supervised object localisation by RL [3, 5, 17, 35], the

proposed IDEAL model requires no accurate object bounding box annotations, therefore more scalable to large size data in practice.
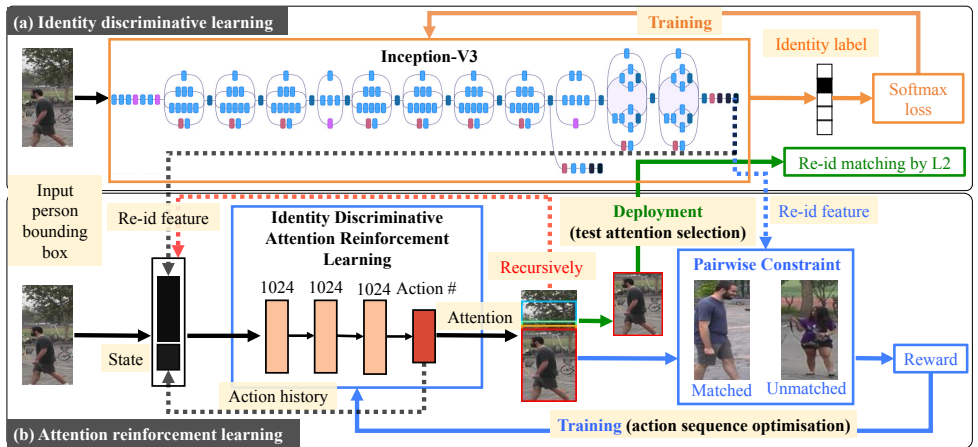

Figure 2: The IDEAL reinforcement learning attention selection model. (a) An identity discriminative learning branch based on the deep Inception-V3 network optimised by a multi-classification softmax loss (orange arrows). (b) An attention reinforcement learning branch designed as a deep Q-network optimised by re-id class label constraints in the deep feature space from branch (a) (blue arrows). For model deployment, the trained attention branch (b) computes the optimal attention regions for each probe and all the gallery images, extract the deep features from these optimal attention regions in the multi-class re-id branch (a) and perform L2 distance matching (green arrows).

3   Re-ID Attention Selection by Reinforcement Learning

The Identity DiscriminativE Attention reinforcement Learning (IDEAL) model has two sub-networks: **(I)** A multi-class *discrimination network* $\mathcal{D}$ by deep learning from a training set of auto-detected person bounding boxes (Fig. 2(a)). This part is flexible with many options from existing deep re-id networks and beyond [11, 52, 54, 62]. **(II)** A re-identification *attention network* $\mathcal{A}$ by reinforcement learning recursively a salient sub-region with its deep feature representation from $\mathcal{D}$ that can maximise identity-matching given re-id label constraints (Fig. 2(b)). Next, we formulate the attention network by reinforcement learning and how this attention network cooperates with the multi-class discrimination network.

3.1   Re-ID Attention Selection Formulation

We formulate the re-id attention selection as a reinforcement learning problem [18]. This allows to correlate directly the re-id attention selection process with the learning objective of an "agent" by recursively *rewarding* or punishing the learning process. In essence, the aim of model learning is to achieve an optimal identity discriminative attending action policy $a = \pi(\mathbf{s})$ of an agent, i.e. a mapping function, that projects a state observation $\mathbf{s}$ (model input) to an action prediction a. In this work, we exploit the Q-learning technique for learning the proposed IDEAL agent, due to its sample efficiency advantage for a small set of actions [16,

(a) Input image   (b) Attending actions (Each red dotted box represents the attention window after the action)

Figure 3: Identity discriminative attending actions are given by an attending scale variable on four directions (left/right/top/bottom). Termination action means the stop of a recursive attending process.

60]. Formally, we aim to learn an optimal state-value function which measures the maximum sum of the current reward (R_t) and all the future rewards ($R_{t+1}, R_{t+2}, \cdots$) discounted by a factor γ at each time step t:

$$Q^*(\mathbf{s}, a) = \max_{\pi} \mathbb{E}\left[R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots \mid \mathbf{s}_t = \mathbf{s}, a_t = a, \pi\right] \qquad (1)$$

Once $Q^*(\mathbf{s}, a)$ is learned, the optimal policy $\pi^*(\mathbf{s})$ can be directly inferred by selecting the action with the maximum $Q^*(\mathbf{s}, a)$ value in model deployment. More specifically, the reinforcement learning agent interacts with each data sample in a sequential episode, which can be considered as a Markov decision process (MDP) [42]. For our purpose, we need to design a specific MDP for re-id discriminative attention selection, as described below.

3.2 Markov Decision Process for Re-ID Attention Selection

We design a MDP for re-id attention selection in auto-detected bounding boxes. In particular, we consider each input person bounding box image as a dynamic environment. An IDEAL agent interacts with this dynamic environment to locate the optimal re-id attention window. To guide this discriminative learning process, we further consider a reward that can encourage those attending actions to improve re-id performance and maximise the cumulative future reward in Eqn. (1). As such, we define actions, states, and rewards as follows.

Actions: An action set **A** is defined to facilitate the IDEAL agent to determine the location and size of an "attention window" (Fig. 3). Specifically, an attending action a is defined by the location shift direction ($a_d \in \{$left, right, top, bottom$\}$) and shift scale ($a_e \in \mathbf{E}$). We also introduce a termination action as a search process stopping signal. **A** consists of a total of $(4 \times |\mathbf{E}| + 1)$ actions. Formally, let the upper-left and bottom-right corner coordinates of the current attention window and an updated attention window be $[x_1, y_1, x_2, y_2]$ and $[x'_1, y'_1, x'_2, y'_2]$ respectively, the action set **A** can then be defined as:

$$\mathbf{A} = \{x'_1 = x_1 + \alpha \Delta x, \ x'_2 = x_2 - \alpha \Delta x, \ y'_1 = y_1 + \alpha \Delta y, \ y'_2 = y_2 - \alpha \Delta y, \ \mathrm{T}\}, \qquad (2)$$
$$\text{where} \quad \alpha \in \mathbf{E}, \quad \Delta x = x_2 - x_1, \quad \Delta y = y_2 - y_1, \quad \mathrm{T} = \text{termination}.$$

Computationally, each action except termination in **A** modifies the environment by cutting off a horizontal or vertical stripe. We set $\mathbf{E} = \{5\%, 10\%, 20\%\}$ by cross-validation in our experiments, resulting in total 13 actions. Such a small attention action space with multi-scale changes has three merits: (1) Only a small number of simple actions are contained, which allows more efficient and stable agent training; (2) Fine-grained actions with small attention changes allow the IDEAL agent sufficient freedoms to utilise small localised regions in auto-detected bounding boxes for subtle identity matching. This enables more effective elimination of undesired background clutter whilst retaining identity discriminative information; (3) The termination action enables the agent to be aware of the satisfactory condition

met for attention selection and stops further actions when optimised.

States: The state $\mathbf{s}_t$ of our MDP at time t is defined as the concatenation of the feature vector $\mathbf{x}_t \in \mathbb{R}^d$ (with d re-id feature dimension) of current attending window and an action history vector $\mathbf{h}_t \in \mathbb{R}^{|\mathbf{E}| \times n_{\text{step}}}$ (with n_{step} a pre-defined maximal action number per bounding box), i.e. $\mathbf{s}_t = [\mathbf{x}_t, \mathbf{h}_t]$. Specifically, at each time step, we extract the feature vector $\mathbf{x}_t$ of current attention window by the trained re-id network $\mathcal{D}$. The action history vector $\mathbf{h}_t$ is a binary vector for keeping a track of all past actions, represented by a $|\mathbf{A}|$-dimensional (13 actions) one-hot vector where the corresponding action bit is encoded as one, all others as zeros.

Rewards: The reward function R (Eqn. (1)) defines the agent task objective. In our context, we therefore correlate directly the reward function of the IDEAL agent's attention behaviour with the re-id matching criterion. Formally, at time step t, suppose the IDEAL agent observes a person image $\mathbf{I}_t$ and then takes an action $a_t = a \in \mathbf{A}$ to attend the image region $\mathbf{I}_t^a$. Given this attention shift from $\mathbf{I}_t$ to $\mathbf{I}_t^a$, its state $\mathbf{s}_t$ changes to $\mathbf{s}_{t+1}$. We need to assess such a state change and signify the agent if this action is encouraged or discouraged by an award or a punishment. To this end, we propose three reward function designs, inspired by pairwise constraint learning principles established in generic information search and person re-id.

Notations From the labelled training data, we sample two other *reference* images w.r.t. $\mathbf{I}_t$: (1) A *cross-view positive* sample $\mathbf{I}_t^+$ sharing the same identity as $\mathbf{I}_t$ but not the camera view; (2) A *same-view negative* sample $\mathbf{I}_t^-$ sharing the camera view as $\mathbf{I}_t$ but not the identity. We compute the features of all these images by $\mathcal{D}$, denoted respectively as $\mathbf{x}_t, \mathbf{x}_t^a, \mathbf{x}_t^+$, and $\mathbf{x}_t^-$.

(I) Reward by Relative Comparison Our first reward function R_t is based on relative comparison, in spirit of the triplet loss for learning to rank [29]. It is formulated as:

$$R_t = R_{rc}(\mathbf{s}_t, a) = \Big(f_{\text{match}}(\mathbf{x}_t^a, \mathbf{x}_t^-) - f_{\text{match}}(\mathbf{x}_t^a, \mathbf{x}_t^+) \Big) - \Big(f_{\text{match}}(\mathbf{x}_t, \mathbf{x}_t^-) - f_{\text{match}}(\mathbf{x}_t, \mathbf{x}_t^+) \Big) \quad (3)$$

where f_{match} defines the re-id matching function. We use the Euclidean distance metric given the Inception-V3 deep features. Intuitively, this reward function commits (i) a positive reward if the attended region becomes more-matched to the *cross-view positive* sample whilst less-matched to the *same-view negative* sample, or (ii) a negative reward otherwise. When a is the termination action, i.e. $\mathbf{x}_t^a = \mathbf{x}_t$, the reward value R_{rc} is set to zero. In this way, the IDEAL agent is supervised to attend the regions subject to optimising jointly two tasks: (1) being more discriminative and/or more salient for the target identity in an inter-view sense (cross-view re-id), whilst (2) pulling the target identity further away from other identities in an intra-view sense (discarding likely shared view-specific background clutter and occlusion therefore focusing more on genuine person appearance). Importantly, this multi-task objective design favourably allows appearance saliency learning to intelligently select the most informative parts of certain appearance styles for enabling holistic clothing patten detection and ultimately more discriminative re-id matching (e.g. Fig. 1(b) and Fig. 4(b)).

(II) Reward by Absolute Comparison Our second reward function considers only the compatibility of a true matching pair, in the spirit of positive verification constraint learning [9]. Formally, this reward is defined as:

$$R_t = R_{ac}(\mathbf{s}_t, a) = \Big(f_{\text{match}}(\mathbf{x}_t, \mathbf{x}_t^+) \Big) - \Big(f_{\text{match}}(\mathbf{x}_t^a, \mathbf{x}_t^+) \Big) \quad (4)$$

The intuition is that, the cross-view matching score of two same-identity images depends on how well irrelevant background clutter/occlusion is removed by the current action. That is, a good attending action will increase a cross-view matching score, and vice verse.

(III) Reward by Ranking Our third reward function concerns the true match ranking change

brought by the agent action, therefore simulating directly the re-id deployment rational [13]. Specifically, we design a binary reward function according to whether the rank of true match $\mathbf{x}_t^+$ is improved when $\mathbf{x}_t$ and $\mathbf{x}_t^a$ are used as the probe separately, as:

$$R_t = R_r(\mathbf{s}_t, a) = \begin{cases} +1, & \text{if } \text{Rank}(\mathbf{x}_t^+|\mathbf{x}_t) > \text{Rank}(\mathbf{x}_t^+|\mathbf{x}_t^a) \\ -1, & \text{otherwise} \end{cases} \tag{5}$$

where $\text{Rank}(\mathbf{x}_t^+|\mathbf{x}_t)$ ($\text{Rank}(\mathbf{x}_t^+|\mathbf{x}_t^a)$) represents the rank of $\mathbf{x}_t^+$ in a gallery against the probe $\mathbf{x}_t$ ($\mathbf{x}_t^a$). Therefore, Eqn. (5) gives support to those actions of leading to a higher rank for the true match, which is precisely the re-id objective. In our implementation, the gallery was constructed by randomly sampling n_g (e.g. 600) cross-view training samples. We evaluate and discuss the above three reward function choices in the experiments (Sec. 4).

3.3   Model Implementation, Training, and Deployment

Implementation and Training For the multi-class discrimination network $\mathcal{D}$ in the IDEAL model, we deploy the Inception-V3 network [48] (Fig. 2(a)), a generic image classification CNN model [48]. It is trained from scratch by a softmax classification loss using person identity labels of the training data. For the re-id attention network $\mathcal{A}$ in the IDEAL model, we design a neural network of 3 fully-connected layers (each with 1024 neurons) and a prediction layer (Fig. 2(b)). This implements the state-value function Eqn. (1). For optimising the sequential actions for re-id attention selection, we utilise the ε-greedy learning algorithm [57] during model training: The agent takes (1) a random action from the action set **A** with the probability ε, and (2) the best action predicted by the agent with the probability $1 - \varepsilon$. We begin with $\varepsilon = 1$ and gradually decrease it by 0.15 every 1 training epoch until reaching 0.1. The purpose is to balance model exploration and exploitation in the training stage so that local minimum can be avoided. To further reduce the correlations between sequential observations, we employ the experience replay strategy [57]. In particular, a fixed-sized memory pool **M** is created to store the agent's N past training sample (experiences) $e_t = (\mathbf{s}_t, a_t, R_t, \mathbf{s}_{t+1})$ at each time step t, i.e. $\{e_{t-N+1}, \cdots, e_t\}$. At iteration i, a mini-batch of training samples is selected randomly from **M** to update the agent parameters θ by the loss function:

$$L_i(\theta_i) = \mathbb{E}_{(\mathbf{s}_t, a_t, R_t, \mathbf{s}_{t+1}) \sim \text{Uniform}(\mathbf{M})} \left(R_t + \gamma \max_{a_{t+1}} Q(\mathbf{s}_{t+1}, a_{t+1}; \tilde{\theta}_i) - Q(\mathbf{s}_t, a_t; \theta_i) \right)^2, \tag{6}$$

where $\tilde{\theta}_i$ are the parameters of an intermediate model for predicting training-time target values, which are updated as θ_i at every ς iterations, but frozen at other times.

Deployment During model deployment, we apply the learned attention network $\mathcal{A}$ to all test probe and gallery bounding boxes for extracting their attention window images. The deep features of these attention window images are used for person re-id matching by extracting the 2,048-D output from the last fully-connected layer of the discrimination network $\mathcal{D}$. We employ the L2 distance as the re-id matching metric.

4   Experiments

Datasets For evaluation, we used two large benchmarking re-id datasets generated by automatic person detection: CUHK03 [23], and Market-1501 [59] (details in Table 1). CUHK03 also provides an extra version of bounding boxes by human labelling therefore offers a like-

to-like comparison between the IDEAL attention selection and human manually cropped images. Example images are shown in (a),(b) and (c) of Fig. 1.

Evaluation Protocol We adopted the standard CUHK03 1260/100 [23] and Market-1501 750/751 [59] training/test person split. We used the single-shot setting on CUHK03, both single- and multi-query setting on Market-1501. We utilised the cumulative matching characteristic (CMC) to measure re-id accuracy. For Market-1501, we also used the recall measure of multiple truth matches by mean Average Precision (mAP).

Implementation Details We implemented the proposed IDEAL method in the TensorFlow framework [0]. We trained an Inception-V3 [48] multi-class identity discrimination network $\mathcal{D}$ from scratch for each re-id dataset at a learning rate of 0.0002 by using the Adam optimiser [19]. The final FC layer output feature vector (2,048-D) together with the L2 distance metric is used as our re-id matching model. All person bounding boxes were resized to 299×299 in pixel. We trained the $\mathcal{D}$ by 100,000 iterations. We optimised the IDEAL attention network $\mathcal{A}$ by the Stochastic Gradient Descent algorithm [4] with the learning rate set to 0.00025. We used the relative comparison based reward function (Eqn. (3)) by default. The experience replay memory (**M**) size for reinforcement learning was 100,000. We fixed the discount factor γ to 0.8 (Eqn. (1)). We allowed a maximum of $n_{step} = 5$ action rounds for each episode in training $\mathcal{A}$. The intermediate regard prediction network was updated every $\varsigma = 100$ iterations. We trained the $\mathcal{A}$ by 10 epochs.

Table 2: Comparing re-id performance. 1st/2nd best results are shown in red/blue. AD: Automatically Detected.

Dataset	CUHK03(AD) [23]				Market-1501(AD) [59]					CUHK03(AD) [23]				Market-1501(AD) [59]			
Metric (%)	R1	R5	R10	R20	Single Query R1	mAP	Multi-Query R1	mAP		R1	R5	R10	R20	Single Query R1	mAP	Multi-Query R1	mAP
ITML[14]	5.1	17.7	28.3	-	-	-	-	-	TMA[29]	-	-	-	-	47.9	22.3	-	-
LMNN[53]	6.3	18.7	29.0	-	-	-	-	-	HL[14]	-	-	-	-	59.5	-	-	-
KISSME[21]	11.7	33.3	48.0	-	40.5	19.0	-	-	HER[52]	60.8	87.0	95.2	97.7	-	-	-	-
MFA[54]	-	-	-	-	45.7	18.2	-	-	FPNN[23]	19.9	-	-	-	-	-	-	-
kLFDA[54]	-	-	-	-	51.4	24.4	52.7	27.4	DCNN+[0]	44.9	76.0	83.5	93.2	-	-	-	-
BoW[59]	23.0	42.4	52.4	64.2	34.4	14.1	42.6	19.5	EDM[44]	52.0	-	-	-	-	-	-	-
XQDA[25]	46.3	78.9	83.5	93.2	43.8	22.2	54.1	28.4	SICI[55]	52.1	84.9	92.4	-	-	-	-	-
MLAPG[26]	51.2	83.6	92.1	96.9	-	-	-	-	SSDAL[47]	-	-	-	-	39.4	19.6	49.0	25.8
L_1-Lap [20]	30.4	-	-	-	-	-	-	-	S-LSTM [51]	57.3	80.1	88.3	-	-	-	61.6	35.3
NFST[56]	53.7	83.1	93.0	94.8	55.4	29.9	68.0	41.9	eSDC[57]	7.7	21.9	35.0	50.0	33.5	13.5	-	-
LSSCDL[58]	51.2	80.8	89.6	-	-	-	-	-	CAN[27]	63.1	82.9	88.2	93.3	48.2	24.4	-	-
SCSP[6]	-	-	-	-	51.9	26.3	-	-	GS-CNN[50]	68.1	88.1	94.6	-	65.8	39.5	76.0	48.4
									IDEAL	71.0	89.8	93.0	95.9	86.7	67.5	91.3	76.2

Comparisons to the State-of-the-Arts We compared the IDEAL model against 24 different contemporary and the state-of-the-art re-id methods (Table 2). It is evident that IDEAL achieves the best re-id performance, outperforming the strongest competitor GS-CNN [50] by 2.9% (71.0-68.1) and 20.9% (86.7-65.8) in Rank-1 on CUHK03 and Market-1501 respectively. This demonstrates a clear positive effect of IDEAL's attention selection on person re-id performance by filtering out bounding box misalignment and random background clutter in auto-detected person images. To give more insight and visualise both the effect of IDEAL and also failure cases, qualitative examples are shown in Fig. 4.

Evaluations on Attention Selection We further compared in more details the IDEAL model against three state-of-the-art saliency/attention based re-id models (eSDC [57], CAN [27], GS-CNN [50]), and two baseline attention methods (Random, Centre) using the Inception-V3 re-id model (Table 3). For *Random Attention*, we attended randomly person bounding boxes by a ratio (%) randomly selected from $\{95, 90, 80, 70, 50\}$. We repeated 10 times and reported the mean results. For *Centre Attention*, we attended all person bounding boxes at

Figure 4: Qualitative evaluations of the IDEAL model: **(a)** Two examples of action sequence for attention selection given by action 1 (Blue), action 2 (Green), action 3 (Yellow), action 4 (Purple), action 5 (Red); **(b)** Two examples of cross-view IDEAL selection for re-id; **(c)** Seven examples of IDEAL selection given by 5, 3, 5, 5, 4, 2, and 2 action steps respectively; **(d)** A failure case when the original auto-detected (AD) bounding box contains two people, manually cropped (MC) gives a more accurate box whilst IDEAL attention selection fails to reduce the distraction; **(e)** Four examples of IDEAL selection on the Market-1501 "distractors" with significantly poorer auto-detected bounding boxes when IDEAL shows greater effects.

Table 3: Comparing attention selection methods. SQ: Single Query; MQ: Multi-Query. AC: Absolute Comparison; RC: Relative Comparison.

Dataset	CUHK03 [23]				Market-1501 [69]			
Metric (%)	R1	R5	R10	R20	R1(SQ)	mAP(SQ)	R1(MQ)	mAP(MQ)
eSDC [67]	7.7	21.9	35.0	50.0	33.5	13.5	-	-
CAN [27]	63.1	82.9	88.2	93.3	48.2	24.4	-	-
GS-CNN [50]	68.1	88.1	**94.6**	-	65.8	39.5	76.0	48.4
No Attention	67.5	88.2	92.6	95.7	84.5	64.8	89.4	72.5
Random Attention	54.1	79.2	85.9	90.4	80.3	54.6	85.1	66.7
Centre Attention (95%)	66.1	86.7	91.1	94.9	84.1	64.2	88.6	69.4
Centre Attention (90%)	64.1	85.3	90.3	93.5	82.7	60.3	87.5	65.3
Centre Attention (80%)	51.9	76.0	83.0	89.0	74.7	48.5	83.4	57.6
Centre Attention (70%)	35.2	62.3	73.2	81.7	63.8	39.0	72.3	43.5
Centre Attention (50%)	16.7	38.8	49.5	62.5	39.9	18.5	46.3	23.9
IDEAL(Ranking)	70.3	89.1	92.7	95.4	86.2	66.3	90.8	74.3
IDEAL(AC)	69.1	88.4	92.1	95.0	85.3	65.5	87.5	72.3
IDEAL(RC)	**71.0**	**89.8**	93.0	**95.9**	**86.7**	**67.5**	**91.3**	**76.2**

centre by one of the same 5 ratios above. It is evident that the IDEAL (Relative Comparison) model is the best. The inferior re-id performance of eSDC, CAN and GS-CNN is due to their strong assumption on accurate bounding boxes. Both Random and Centre Attention methods do not work either with even poorer re-id accuracy than that with "No Attention" selection. This demonstrates that optimal attention selection given by IDEAL is non-trivial. Among the three attention reward functions, *Absolute Comparison* is the weakest, likely due to the lack of reference comparison against false matches, i.e. no population-wise matching context in attention learning. *Ranking* fares better, as it considers reference comparisons. The extra advantage of *Relative Comparison* is due to the *same-view negative* comparison in Eqn.(3). This provides a more reliable background clutter detection since same-view images

are more likely to share similar background patterns.

Auto-Detection+IDEAL vs. Manually Cropped Table 4 shows that auto-detection+IDEAL can perform similarly to that of *manually cropped* images in CUHK03 test[1], e.g. 71.0% vs. 71.9% for Rank-1 score. This shows the potential of IDEAL in eliminating expensive manual labelling of bounding boxes and for scaling up re-id to large data deployment.

Table 4: Auto-detection+IDEAL vs. manually cropped re-id on CUHK03.

Metric (%)	R1	R5	R10	R20
Auto-Detected+**IDEAL**	71.0	89.8	93.0	95.9
Manually Cropped	**71.9**	**90.4**	**94.5**	**97.1**

Effect of Action Design We examined three designs with distinct attention scales. Table 5 shows that the most fine-grained design $\{5\%, 10\%, 20\%\}$ is the best. This suggests that the re-id by appearance is subtle and small regions make a difference in discriminative matching.

Table 5: Attention action design evaluation. SQ: Single Query; MQ: Multi-Query.

Dataset	CUHK03 [23]				Market-1501 [69]			
Metric (%)	R1	R5	R10	R20	R1(SQ)	mAP(SQ)	R1 (MQ)	mAP(MQ)
$\{5\%, 10\%, 20\%\}$	**71.0**	**89.8**	**93.0**	**95.9**	**86.7**	**67.5**	**91.3**	**76.2**
$\{10\%, 20\%, 30\%\}$	68.3	88.1	91.8	95.0	86.2	66.8	90.5	73.4
$\{10\%, 20\%, 50\%\}$	67.6	87.5	91.4	93.9	85.3	65.6	88.8	72.1

5   Conclusion

We presented an Identity DiscriminativE Attention reinforcement Learning (IDEAL) model for optimising re-id attention selection in auto-detected bounding boxes. This improves notably person re-id accuracy in a fully automated process required in practical deployments. The IDEAL model is formulated as a unified framework of discriminative identity learning by a deep multi-class discrimination network and attention reinforcement learning by a deep Q-network. This achieves jointly optimal identity sensitive attention selection and re-id matching performance by a reward function subject to identity label pairwise constraints. Extensive comparative evaluations on two auto-detected re-id benchmarks show clearly the advantages and superiority of this IDEAL model in coping with bounding box misalignment and background clutter removal when compared to the state-of-the-art saliency/attention based re-id models. Moreover, this IDEAL automatic attention selection mechanism comes near to be equal to human manual labelling of person bounding boxes on re-id accuracy, therefore showing a great potential for scaling up automatic re-id to large data deployment.

6   Acknowledgements

---

[1]The Market-1501 dataset provides no manually cropped person bounding boxes.

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv*, 2016.

[2] Ejaz Ahmed, Michael J. Jones, and Tim K. Marks. An improved deep learning architecture for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[3] Miriam Bellver, Xavier Giró-i Nieto, Ferran Marqués, and Jordi Torres. Hierarchical object detection with deep reinforcement learning. *arXiv preprint arXiv:1611.03718*, 2016.

[4] Léon Bottou. Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pages 421–436. 2012.

[5] Juan C Caicedo and Svetlana Lazebnik. Active object localization with deep reinforcement learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2488–2496, 2015.

[6] Dapeng Chen, Zejian Yuan, Badong Chen, and Nanning Zheng. Similarity learning with spatial constraints for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[7] Ying-Cong Chen, Xiatian Zhu, Wei-Shi Zheng, and Jian-Huang Lai. Person re-identification by camera correlation aware feature augmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PP(99):1–1, 2017.

[8] Dong Seon Cheng, Marco Cristani, Michele Stoppa, Loris Bazzani, and Vittorio Murino. Custom pictorial structures for re-identification. In *British Machine Vision Conference*, 2011.

[9] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[10] Jason V. Davis, Brian Kulis, Prateek Jain, Suvrit Sra, and Inderjit S. Dhillon. Information-theoretic metric learning. In *International Conference on Machine learning*, 2007.

[11] Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao. Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition*, 48(10):2993–3003, 2015.

[12] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, 2010.

[13] Shaogang Gong, Marco Cristani, Change Loy Chen, and Timothy M. Hospedales. The re-identification challenge. In *Person Re-Identification*. Springer, 2014.

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

[15] Douglas Gray, Shane Brennan, and Hai Tao. Evaluating appearance models for recognition, reacquisition and tracking. In *IEEE International Workshop on Performance Evaluation for Tracking and Surveillance*, 2007.

[16] Shixiang Gu, Timothy Lillicrap, Zoubin Ghahramani, Richard E Turner, and Sergey Levine. Q-prop: Sample-efficient policy gradient with an off-policy critic. 2017.

[17] Zequn Jie, Xiaodan Liang, Jiashi Feng, Xiaojie Jin, Wen Lu, and Shuicheng Yan. Tree-structured reinforcement learning for sequential object localization. In *Advances in Neural Information Processing Systems*, pages 127–135, 2016.

[18] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.

[19] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv*, 2014.

[20] Elyor Kodirov, Tao Xiang, Zhenyong Fu, and Shaogang Gong. Person re-identification by unsupervised l1 graph learning. In *European Conference on Computer Vision*, 2016.

[21] Martin Koestinger, Martin Hirzer, Paul Wohlhart, Peter M. Roth, and Horst Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 2012.

[23] Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang. Deepreid: Deep filter pairing neural network for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[24] Wei Li, Xiatian Zhu, and Shaogang Gong. Person re-identification by deep joint learning of multi-loss classification. In *International Joint Conference of Artificial Intelligence*, 2017.

[25] Shengcai Liao and Stan Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In *IEEE International Conference on Computer Vision*, 2015.

[26] 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*, 2015.

[27] Hao Liu, Jiashi Feng, Meibin Qi, Jianguo Jiang, and Shuicheng Yan. End-to-end comparative attention networks for person re-identification. *arXiv*, 2016.

[28] Siqi Liu, Zhenhai Zhu, Ning Ye, Sergio Guadarrama, and Kevin Murphy. Optimization of image description metrics using policy gradient methods. *arXiv:1612.00370*, 2016.

[29] Tie-Yan Liu et al. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*, 3(3):225–331, 2009.

[30] Chen Change Loy, Tao Xiang, and Shaogang Gong. Multi-camera activity correlation analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[31] Chen Change Loy, Chunxiao Liu, and Shaogang Gong. Person re-identification by manifold ranking. In *IEEE International Conference on Image Processing*, 2013.

[32] Xiaolong Ma, Xiatian Zhu, Shaogang Gong, Xudong Xie, Jianming Hu, Kin-Man Lam, and Yisheng Zhong. Person re-identification by unsupervised video matching. *Pattern Recognition*, 65:197–210, 2017.

[33] Mohsen Malmir, Karan Sikka, Deborah Forster, Ian Fasel, Javier R Movellan, and Garrison W Cottrell. Deep active object recognition by joint label and action prediction. *Computer Vision and Image Understanding*, 156:128–137, 2017.

[34] Niki Martinel, Abir Das, Christian Micheloni, and Amit K Roy-Chowdhury. Temporal model adaptation for person re-identification. In *European Conference on Computer Vision*, 2016.

[35] Stefan Mathe, Aleksis Pirinen, and Cristian Sminchisescu. Reinforcement learning for visual object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2894–2902, 2016.

[36] Alexis Mignon and Frédéric Jurie. Pcca: A new approach for distance learning from sparse pairwise constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

[37] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518 (7540):529–533, 2015.

[38] Sakrapee Paisitkriangkrai, Chunhua Shen, and Anton van den Hengel. Learning to rank in person re-identification with metric ensembles. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[39] Sateesh Pedagadi, James Orwell, Sergio A. Velastin, and Boghos A. Boghossian. Local fisher discriminant analysis for pedestrian re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[40] Peixi Peng, Yonghong Tian, Tao Xiang, Yaowei Wang, Massimiliano Pontil, and Tiejun Huang. Joint semantic and latent attribute modelling for cross-class transfer learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[41] Bryan Prosser, Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Person re-identification by support vector ranking. In *British Machine Vision Conference*, 2010.

[42] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994. ISBN 0471619779.

[43] Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jarret Ross, and Vaibhava Goel. Self-critical sequence training for image captioning. *arXiv:1612.00563*, 2016.

[44] Yang Shen, Weiyao Lin, Junchi Yan, Mingliang Xu, Jianxin Wu, and Jingdong Wang. Person re-identification with correspondence structure learning. In *IEEE International Conference on Computer Vision*, pages 3200–3208, 2015.

[45] Hailin Shi, Xiangyu Zhu, Shengcai Liao, Zhen Lei, Yang Yang, and Stan Z Li. Constrained deep metric learning for person re-identification. *arXiv preprint arXiv:1511.07545*, 2015.

[46] Hailin Shi, Yang Yang, Xiangyu Zhu, Shengcai Liao, Zhen Lei, Weishi Zheng, and Stan Z Li. Embedding deep metric for person re-identification: A study against large variations. In *European Conference on Computer Vision*, 2016.

[47] Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Deep attributes driven multi-camera person re-identification. In *European Conference on Computer Vision*, pages 475–491. Springer, 2016.

[48] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[49] Evgeniya Ustinova and Victor Lempitsky. Learning deep embeddings with histogram loss. In *Advances in Neural Information Processing Systems*, pages 4170–4178, 2016.

[50] Rahul Rama Varior, Mrinal Haloi, and Gang Wang. Gated siamese convolutional neural network architecture for human re-identification. In *European Conference on Computer Vision*, 2016.

[51] Rahul Rama Varior, Bing Shuai, Jiwen Lu, Dong Xu, and Gang Wang. A siamese long short-term memory architecture for human re-identification. In *European Conference on Computer Vision*, 2016.

[52] Faqiang Wang, Wangmeng Zuo, Liang Lin, David Zhang, and Lei Zhang. Joint learning of single-image and cross-image representations for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[53] Hanxiao Wang, Shaogang Gong, and Tao Xiang. Unsupervised learning of generative topic saliency for person re-identification. In *British Machine Vision Conference*, Nottingham, United Kingdom, September 2014.

[54] Hanxiao Wang, Shaogang Gong, and Tao Xiang. Highly efficient regression for scalable person re-identification. In *British Machine Vision Conference*, 2016.

[55] Hanxiao Wang, Shaogang Gong, Xiatian Zhu, and Tao Xiang. Human-in-the-loop person re-identification. In *European Conference on Computer Vision*, 2016.

[56] Hanxiao Wang, Xiatian Zhu, Tao Xiang, and Shaogang Gong. Towards unsupervised open-set person re-identification. In *IEEE International Conference on Image Processing*, 2016.

[57] T. Wang, S. Gong, X. Zhu, and S. Wang. Person re-identification by discriminative selection in video ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, January 2016.

[58] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by video ranking. In *European Conference on Computer Vision*, 2014.

[59] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by discriminative selection in video ranking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(12):2501–2514, 2016.

[60] Christopher John Cornish Hellaby Watkins. *Learning from delayed rewards*. PhD thesis, University of Cambridge England, 1989.

[61] Kilian Q. Weinberger and Lawrence K. Saul. Distance metric learning for large margin nearest neighbor classification. *The Journal of Machine Learning Research*, 10:207–244, December 2009.

[62] Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang. Learning deep feature representations with domain guided dropout for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[63] Tong Xiao, Shuang Li, Bochao Wang, Liang Lin, and Xiaogang Wang. End-to-end deep learning for person search. *arXiv:1604.01850*, 2016.

[64] Fei Xiong, Mengran Gou, Octavia Camps, and Mario Sznaier. Person re-identification using kernel-based metric learning methods. In *European Conference on Computer Vision*. 2014.

[65] Li Zhang, Tao Xiang, and Shaogang Gong. Learning a discriminative null space for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[66] Ying Zhang, Baohua Li, Huchuan Lu, Atshushi Irie, and Xiang Ruan. Sample-specific svm learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[67] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Unsupervised salience learning for person re-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[68] Rui Zhao, Wanli Ouyang, and Xiaogang Wang. Person re-identification by salience matching. In *IEEE International Conference on Computer Vision*, 2013.

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

[70] Liang Zheng, Hengheng Zhang, Shaoyan Sun, Manmohan Chandraker, and Qi Tian. Person re-identification in the wild. *arXiv preprint arXiv:1604.02531*, 2016.

[71] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Re-identification by relative distance comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 653–668, March 2013.

[72] Wei-Shi Zheng, Xiang Li, Tao Xiang, Shengcai Liao, Jianhuang Lai, and Shaogang Gong. Partial person re-identification. In *IEEE International Conference on Computer Vision*, pages 4678–4686, 2015.