Vehicle Re-Identification by Fine-Grained Cross-Level Deep Learning

Aytaç Kanacı[1]
a.kanaci@qmul.ac.uk
Xiatian Zhu[2]
eddy@visionsemantics.com
Shaogang Gong[1]
s.gong@qmul.ac.uk

[1] Computer Vision Group, EECS, Queen Mary University of London, London E1 4NS, UK
[2] Vision Semantics Ltd., London E1 4NS, UK

Abstract

Vehicle re-identification in unconstrained images is a challenging computer vision task due to the subtle visual appearance discrepancy between different vehicle identities and large visual appearance changes of the same vehicle instance in different camera views with uncontrolled illumination, view-angle, low-resolution, and background clutters. Existing methods often rely heavily on the availability of cross-camera identity pairwise annotations collected by exhaustive human labelling. This approach is unscalable to many real-world deployment scenarios with limited access to both labelling budgets and vehicle re-appearance between every camera pair. In this work, we solve these challenges by exploiting the inherent hierarchical structure information of vehicle identity and vehicle model class so to eliminate the need for identity level label collection. Specifically, we propose to transfer the vehicle model discriminative representation for more fine-grained re-id tasks by fully leveraging the strong capacity of existing deep models in learning cross-level representations. This realises "Cross-Level Vehicle Recognition" (CLVR). Extensive comparative experiments demonstrate the superiority of the proposed CLVR method over state-of-the-art approaches to using fine-grained identity pairwise labels on the largest vehicle re-id benchmarking dataset.

1 Introduction

Vehicle re-identification (re-id) aims to match the identity of vehicle bounding box images across non-overlapping camera views deployed over open-world surveillance spaces. This is an inherently challenging task because vehicle visual appearance may vary dramatically in different view angles captured by distinct camera views with unknown covariates in illumination, occlusion, and background clutter [15]. Compared to open space person re-identification [6], vehicle re-id is largely under-studied but critical for intelligent transport and forensic analysis [10] in creating modern smart cities across the world. Motivated by the extensive works on person re-id and the capacity of deep models for learning from large sized training data, recent vehicle re-id methods are typically designed to learn an identity discriminative deep feature representation [15, 18]. This requires the access to a very large cross-camera identity *pairwise* training dataset acquired by costly and time-consuming

Figure 1: Illustration of vehicle re-identification challenges: **(Top)**: Stark visual similarities between different vehicle instances of the same model class; **(Bottom)**: Significant variations in illumination, view-angle, and background clutters.

human labelling. They are therefore not scalable to many real-world application scenarios when: (1) No sufficient manual labelling budget is available; (2) There may not exist a sufficiently large number of training vehicles reappearing in every pair of camera views.

One question that is never asked: How important is identity pairwise cross-views labelling for vehicle re-id modelling as compared to vehicle model labelling? This is because: **(1)** Relative to the former, the latter labelling is much *easier to collect* without the need for cross-camera vehicle reappearing, although less fine-grained with weaker supervision information. **(2)** Vehicle identity is intrinsically associated with the model classes, e.g. two vehicle image instances of the same identity *must share the same model class*, but the inverse does not necessarily hold. In other words, the model and identity labels form a top-down hierarchical structure. **(3)** There exists hundreds of different vehicle models with merely small visual appearance differences between some model classes. This means that model class labels are already very *fine-grained* and potentially provide notably discriminative information relevant to vehicle re-id tasks. **(4)** An individual vehicle identity is more fine-grained than its class category with potentially almost *indistinguishable differences* among different identities (Figure 1). In this work, we investigate the usefulness of fine-grained vehicle model detection for even more fine-grained vehicle instance search and re-identification without the need of cross-camera vehicle identity pairwise instance labelling for model training.

The **contributions** of this work are: **(1)** We propose a vehicle discriminative learning model for more fine-grained vehicle instance re-identification task so that expensive and time-consuming cross-camera identity pairwise labelling can be avoided. We call this method "Cross-Level Vehicle Recognition" (CLVR). This cross-level matching scheme is significantly different from existing methods that typically rely on the availability of identity instance annotations for model discrimination acquisition, due to only relatively coarser and cheaper vehicle model annotations are needed. Apart from reducing labelling cost, this approach takes into account that vehicle identity instance labelling may be over fine-grained and can potentially impose negative impact to re-id model optimisation due to the strong similarities of different instances of the same vehicle model. To our best knowledge, this is the first attempt of exploiting the potentials of vehicle model information for semantically correlated instance level re-identification tasks. **(2)** We present a simple but effective CLVR instantiation model for vehicle re-identification by exploiting state-of-the-art deep Convolutional Neural Network (CNN) models (e.g. Inception-V3 [26]) for achieving not only accurate vehicle model classification but also reliable vehicle instance re-identification

beyond the less fine-grained model-level recognition. Extensive comparative evaluations demonstrate the superiority of the proposed CLVR method over existing state-of-the-art vehicle re-id models (Coupled Cluster Loss [15] and MixedDiff [15]) on the largest vehicle re-id benchmarking dataset.

2 Related Work

Vehicle Model Classification One closely related problem to re-identification is vehicle model classification [7, 10, 13, 14, 25, 33]. But, the two problems are usually studied independently. For example, Yang et al. [33] propose a part attributes driven vehicle model recognition. They also contribute a large comprehensive car dataset named "CompCars" with model class labels but without vehicle identity labels. More recently, Hu et al. [10] formulate a deep CNN framework capable of selecting spatial salient vehicle parts in order to learn more discriminative model representations without explicit parts annotations, with two model classification datasets CarFlag-563 and CarFlag-1532 introduced. Different from these existing works, we uniquely exploit the easily available vehicle model labels for more fine-grained re-id tasks, i.e. cross-level vehicle visual analysis. As a result, our approach allows benefiting automatically from the new developments in this research line.

Vehicle Re-Identification Compared to person re-identification by either faces [8, 11, 16, 20, 27, 34] or whole bodies [3, 5, 6, 12, 19, 21, 28, 29, 30, 31, 32, 35], vehicle re-id is significantly under-studied. Recent works [15, 17, 18] are mostly introductions of new datasets and benchmarking with standard deep CNN model results on those datasets. For vehicle re-id model optimisation, the existing methods typically require a very large number of cross-view identity pairwise annotations *in addition to* the model class labels. For example, the current state-of-the-art MixedDiff [15] method modifies the conventional triplet loss [23] with a class-level representation in the Siamese CNN framework and trains the deep model with both vehicle model and identity labels in a complex multi-stages process. This significantly restricts the transferability of these strongly supervised methods due to the extremely high labelling cost of collecting cross-view pairwise identity instance annotations. This type of approaches is largely motivated by the extensive person re-id methods [6] due to their similar nature in the problem level.

However, vehicles are uniquely different from people in their appearances due to their shared structures in manufacturing: (1) the vehicle model category and/or production year (labels that are less fine-grained then identity labels) when correctly classified or recognized are also informative of the identity label because of the hierarchical nature of these labels (2) vehicles of the same model that are the same colour are visually identical as shown in Figure 1. This difference has not been exploited in the existing vehicle re-id methods.

In contrast to all these existing methods, we uniquely bridge the connection between vehicle model classification and vehicle re-id, by investigating the discrimination capability of vehicle model sensitive deep features in performing more fine-grained identity matching tasks. To our best knowledge, this is the first systematic attempt of investigating this structural knowledge inherent to man-made vehicles for scaling up vehicle re-id modelling by proposing a cross-level vehicle recognition approach in the hope of eliminating the tedious identity-level fine-grained labelling requirement.

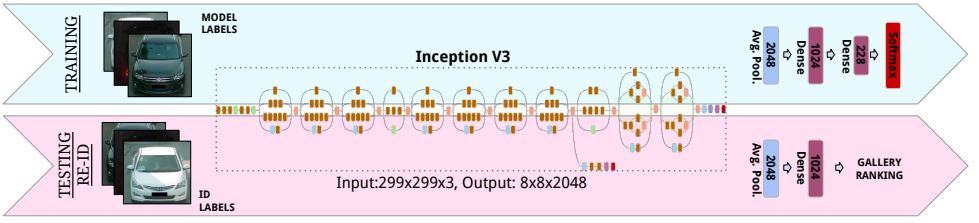

Figure 2: Overview of the proposed Cross-Level Vehicle Recognition (CLVR) method for vehicle re-identification: **(1) Training** (vehicle model classification): Learn a less fine-grained vehicle model classification deep model by a customised Inception-V3 [26] CNN network; **(2) Deployment** (vehicle re-id matching): Deploy the learned CLVR model as a feature extractor using the output of the fully-connected feature (Dense-1024) layer for more fine-grained vehicle re-id tasks.

3 Cross-Level Vehicle Recognition Modelling

3.1 Problem Statement

We aim to learn a deep representation model for a generic distance matching (e.g. L2) based vehicle re-identification without the need for tedious identity labels in model training, instead only less fine-grained vehicle model labels are exploited. We assume a set of n vehicle bounding box training images $\mathcal{I} = \{\boldsymbol{I}_i\}_{i=1}^n$ with the corresponding vehicle model class labels as $\mathcal{Y} = \{y_i\}_{i=1}^n$. These training images capture the visual appearance and variation of n_{model} (where $y_i \in [1, \cdots, n_{\text{model}}]$) different vehicle model categories under multiple non-overlapping camera views. A model needs to learn from these image-model correspondence relations and critically, to transfer the learned knowledge to recognise other unseen vehicle identity instances (more fine-grained) in model deployment. We call this method "Cross-Level Vehicle Recognition". This is in contrast to most existing vehicle re-id methods typically depending only on learning from cross-camera identity pairwise labels. In comparison, vehicle model labels are much cheaper to annotate than cross-view vehicle instance pairwise labelling due to no need of searching cross-view re-appearance of the same vehicle identity.

3.2 Learning Vehicle Model Deep CNN Architecture

The overall network design of the proposed CLVR method is depicted in Figure 2. Specifically, we construct the CLVR by customising the 42-layers Inception-V3 CNN architecture design [26] due to its high computational cost-efficiency (higher modelling capacity at a smaller parameter size) and the capability for learning more discriminative visual features at varying spacial scales. We modify the network by (1) removing the original 1000-D classification layer (for ImageNet 1,000 class) and (2) adding a fully-connected *feature* layer with 1024 neurons on top of the Inception-V3 average pooling layer, followed by a new *classification* layer for accommodating the 228 vehicle model classes. Other competitive architectures, e.g. ResNet [9] or VGG-Net [24], can be modified in a similar manner for this purpose. For model training, we utilise the Softmax classification loss function to optimise vehicle model discrimination given training labels of multiple classes. Formally, we predict

the posterior probability $\tilde{y}_i$ of training image $\boldsymbol{I}_i$ over the given vehicle model label y_i:

$$p_i = p(\tilde{y}_i = y_i|\boldsymbol{I}_i) = \frac{\exp(\boldsymbol{w}_{y_i}^\top \boldsymbol{x}_i)}{\sum_{k=1}^{n_{\text{model}}} \exp(\boldsymbol{w}_k^\top \boldsymbol{x}_i)} \tag{1}$$

where $\boldsymbol{x}_i$ refers to the feature vector of $\boldsymbol{I}_i$ from the CLVR CNN model, and $\boldsymbol{w}_k$ the prediction function parameter of training model class k. The model training loss on a mini-batch of n_{bs} images is computed as:

$$l = -\frac{1}{n_{\text{bs}}} \sum_{i=1}^{n_{\text{bs}}} \log\Big(p(\tilde{y}_i = y_i|\boldsymbol{I}_i)\Big) \tag{2}$$

3.3 Vehicle Re-ID by Cross-Level Vehicle Model Representation

After the CLVR deep CNN model is trained with vehicle model class annotations, we deploy the last fully connected layer output (1024-D vector) as feature representation for more fine-grained vehicle re-id at the instance level. We utilise *only* a generic distance metric *without* camera-pair specific distance metric learning, e.g. L2 distance. Specifically, given a test probe vehicle image $\boldsymbol{I}^p$ from one camera view and a set of test gallery images $\{\boldsymbol{I}_i^g\}$ from other non-overlapping camera views: (1) We first compute their corresponding 1024-D feature vectors by forward-feeding vehicle images into the trained CLVR model, denoted as $\boldsymbol{x}^p$ and $\{\boldsymbol{x}_i^g\}$. (2) We then compute the cross-camera matching score between $\boldsymbol{x}^p$ and $\boldsymbol{x}_i^g$ by L2 distance. (3) We lastly rank all gallery images in ascending order by their matching distances to the probe image. The probabilities of true matches of probe person images in Rank-1 and among the higher ranks indicate the goodness of the learned CLVR deep features for vehicle re-id tasks.

4 Experiments

Dataset For evaluation, we selected the recent large vehicle re-identification dataset VehicleID [15]. This dataset provides a standard training/test images split: (1) 113,346 images of 13,164 identities for model training (8.61 images per identity); and (2) non-overlapping 108,211 image of 13,164 identities for test evaluation (8.22 images per identity). Of which 90,168 images in the training set are also labelled with vehicle model categories. Note that, only vehicle model labels are required for training the proposed CLVR deep model. In total, there are 228 vehicle model classes, with many classes presenting only very subtle visual differences. This causes the typical fine-grained recognition challenges, further compounded by the uncontrolled appearance variations in illumination, pose, view-angle, and background clutters (see examples in Figure 3).

Evaluation Protocol We adopted the benchmarking setting of [15]. There are three different sets of vehicle images for testing re-id: *small* (6,493 images of 800 identities), *medium* (13,377 images of 1,600 identities), and *large* (19,777 images of 2,400 identities). For each case, one image per identity is randomly selected from the gallery set as the probe image, whilst the remaining images are put into the gallery set. To train our CLVR (vehicle model classification), the images with model labels are divided into a random 80%/20% training/test split per class. We ended up with 72,049 training images and 18,119 test images for vehicle model classification. For performance evaluation, we used the cumulative matching characteristic (CMC) as vehicle re-identification performance measure [15]. The

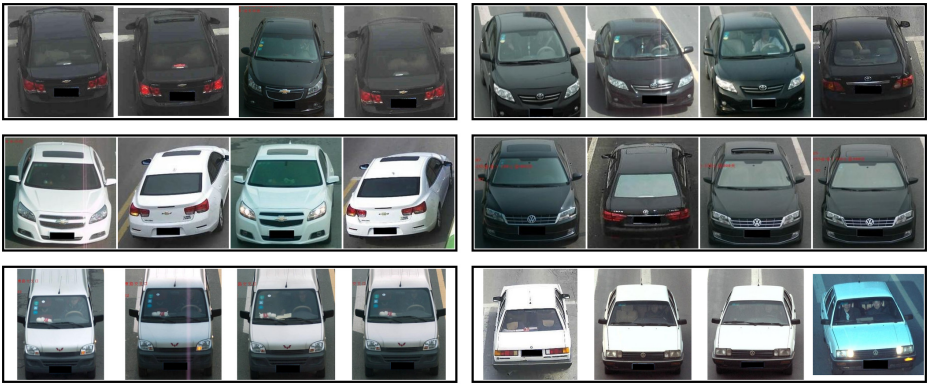

Figure 3: Example vehicle images from the VehicleID dataset [15]. All four images in each group describe the same vehicle identity under different imaging conditions. It is evident that: (1) Vehicle images of the same identity may have large visual appearance differences; (2) Whilst vehicle images of different identities may appear very similarly.

CMC is computed on each individual rank position k as the probe cumulative percentage of truth matches appearing at ranks $\leq k$. For vehicle model classification, the common accuracy measure is used [4].

Implementation Details We implemented the proposed vehicle re-id model in the Tensorflow [1] framework. For model learning, we took the canonical multi-stage sequential training strategy. In the first warm-up stage, we firstly initialised the Inception-V3 [26] network weights with the ImageNet-1K object class images [22]. We then modified the Inception-V3 architecture by removing the last 1000-D classification full-connection layer and added randomly initialised Dense layer at the end. To initialising these new layers, we froze all ImageNet-1K trained layers (indicated by the dotted rectangle in Figure 2) and trained the network by 10 epochs using the RMSprop optimiser at learning rate 0.001. In the second target training phase, the network was trained additional 30 epochs using Stochastic Gradient Descent (SGD) optimiser at learning rate 0.0001. All input images are resized into 299×299 in pixel. The mini-batch size n_{bs} is set to 100.

4.1 Evaluations on Vehicle Model Classification

In this section, we evaluated the CLVR generalisation capability on recognising less fine-grained vehicle model classes. Overall, the proposed CLVR model achieves 94.8% vehicle model classification accuracy over all 228 model classes. This suggests the satisfactory performance of our learned deep features in distinguishing the subtle visual discrepancy between different but very similar vehicle model classes. We further examined the per-class recognition performance. Figure 4 (Left) shows that the vast majority classes can be very accurately recognised, whilst a few obtains very low (even 0%) accuracy. This is mainly because of only very sparse corresponding training images available for these poorly detected classes (Figure 4 (Right)). For visual evaluation, we show some vehicle model recognition examples in Figure 5.

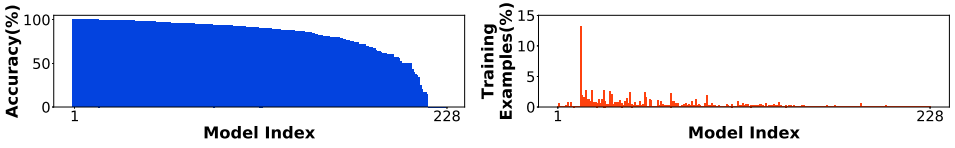

Figure 4: **(Left)**: Classification accuracies over all 228 vehicle model classes; **(Right)**: The training image size distribution over the corresponding model classes of (Left).

Figure 5: Qualitative evaluations of vehicle model classification. **(Left)**: Correctly classified vehicle images with large visual appearance similarity between different model classes. **(Right)**: Misclassified vehicle images due to extreme illumination conditions.

4.2 Evaluations on Cross-Level Vehicle Re-Identification

We evaluated the performance of the CLVR method in vehicle re-id tasks by using the less fine-grained vehicle model deep features for directly performing the re-id tasks on test images with identity labels.

Competitors We compared the proposed approach with the following two state-of-the-art deep methods: **(1)** Coupled Cluster Loss (CCL) [15]: A variant of triplet loss which uniquely replaces the anchor samples with the corresponding class centre for that mini-batch in order to suppress the negative effects of improper anchors in batch-wise model optimisation. **(2)** MixedDiff [15]: A mixture design of the CCL and Softmax classification loss functions by exploiting both vehicle model and identity labels in a two branch deepp neural network model. One branch is trained with the vehicle model labels using the Softmax classification loss and the other with identity labels using the CCL triplet loss. Both deep features are then concatenated and fed through a two fully-connected layers subnetwork optimised using another CCL loss. Consequently, the whole training process consists of three different stages each with the need for careful parameter tuning. The base network structure for MixedDiff model is VGG-M [2]

Comparative Results We present the comparative results in Table 1. The proposed CLVR method significantly outperforms the state-of-the-art methods, e.g. surpassing CCL and MixedDiff by 18.4% (62.0-43.6) and 13.0% (62.0-49.0) in rank-1 rate, respectively. Note that, the competitors achieve their performance by utilising additional more fine-grained identity pairwise supervision with a more complex deep model training process. This is in contrast to the CLVR design of exploiting only the cheaper vehicle model annotation in a

Table 1: Vehicle re-identification performance comparisons. Metric: CMC measure (%).

Method	Pairwise	Label Type	Rank	Small (800)	Medium (1600)	Large (2400)
CCL[15]	✓	ID	1	43.6	37.0	32.9
MixedDiff[15]	✓	Model&ID	1	49.0	42.8	38.2
CLVR	✗	Model	1	**62.0**	**56.1**	**50.6**
CCL [15]	✓	ID	5	64.2	57.1	53.3
MixedDiff[15]	✓	Model&ID	5	73.5	66.8	61.6
CLVR	✗	Model	5	**76.0**	**71.8**	**68.0**

much simpler way of training, e.g. standard random sampling for constructing mini-batches with no complicated pairwise data sampling such as hard-negative triplet mining [23]. These evidences suggest that vehicle model classes supervised deep features can be very effective and useful for cross-level vehicle re-id if exploited properly. While expensive identity pairwise labels offer more fine-grained information, model optimisation is likely to get confused simultaneously due to the over subtle and possibly no distinguishable supporting visual evidences in training image data. The performance advantages of CLVR over the alternatives remain on Rank-5 rate in cases of different gallery search space sizes, although lower than that of Rank-1. This is because vehicle model label is not sufficiently informative for distinguishing those different identities of the same model category. We further show visual examples of vehicle re-id tasks in Figure 6.

Figure 6: Qualitative evaluations of vehicle re-id by the proposed CLVR method. In each panel (Left/Middle/Right), one row shows a specific re-id task: In the 1st column is the probe image, followed by the list of ranks 1∼5 of the gallery images. True and false gallery matches of the probe image are indicated by green and red boxes, respectively. **(Left)**: The rank-1 images are true matches. **(Middle)**: A true match among ranks 1∼5. **(Right)**: All ranks 1∼5 are false matches. It is worth noting that all top-5 gallery matches have the same viewpoint as the probe vehicle in all cases. This suggests that the CLVR model is able to learn highly view-sensitive vehicle fine-grained features from the training images and the cheaper model labelled automatically.

5 Conclusion

In this work, we present a Cross-Level Vehicle Recognition (CLVR) method for addressing the vehicle re-identification problem by flexibly exploiting the hierarchical knowledge structure inherent to vehicle identity and vehicle model classes. This favourably avoids the labour-intensive requirement of very large cross-view identity pairwise training data for model learning by existing alternative approaches. Specifically, we propose to transfer the vehicle model discriminative feature representations for more fine-grained vehicle identity matching by exploiting fully the strong capacity of state-of-the-art deep models in learning cross-level transferable features. This cross-level deep learning scheme brings with not only a much lower cost in collecting model training data, but also a simpler model training requirement without notorious experience driven mini-batch training data construction tricks. We have validated the superiority and advantages of the proposed CLVR method over state-of-the-art deep learning alternatives by extensive comparative evaluations on the largest vehicle re-identification benchmark dataset. We also provide a number of qualitative evaluations for offering visual insights.

Acknowledgements

This work was partially supported by DeepInsight (Deep Learning for Large Scale Video Analysis), Vision Semantics Ltd., and Royal Society Newton Advanced Fellowship Programme (NA150459).

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] Ken Chatfield, Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *British Machine Vision Conference*, 2015.

[3] 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.

[4] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

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

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

[7] Hui-Zhen Gu and Suh-Yin Lee. Car model recognition by utilizing symmetric property to overcome severe pose variation. *Machine Vision and Applications*, 24(2):255–274, 2013.

[8] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[10] Qichang Hu, Huibing Wang, Teng Li, and Chunhua Shen. Deep cnns with spatially weighted pooling for fine-grained car recognition. *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[11] Gary B Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[12] 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.

[13] Liang Liao, Ruimin Hu, Jun Xiao, Qi Wang, Jing Xiao, and Jun Chen. Exploiting effects of parts in fine-grained categorization of vehicles. In *IEEE International Conference on Image Processing*, 2015.

[14] Yen-Liang Lin, Vlad I Morariu, Winston Hsu, and Larry S Davis. Jointly optimizing 3d model fitting and fine-grained classification. In *European Conference on Computer Vision*, 2014.

[15] Hongye Liu, YongHong Tian, Yaowei Wang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[16] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[17] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. In *IEEE International Conference on Multimedia and Expo*, 2016.

[18] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *European Conference on Computer Vision*, 2016.

[19] 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.

[20] Aaron Nech and Ira Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[21] 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.

[22] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115 (3):211–252, 2015.

[23] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[24] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[25] Jakub Sochor, Adam Herout, and Jiri Havel. Boxcars: 3d boxes as cnn input for improved fine-grained vehicle recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3006–3015, 2016.

[26] 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*.

[27] Luan Tran, Xi Yin, and Xiaoming Liu. Disentangled representation learning gan for pose-invariant face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

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

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

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

[31] Taiqing Wang, Shaogang Gong, Xiatian Zhu, and Shengjin Wang. Person re-identification by video ranking. In *European Conference on Computer Vision*, pages 688–703, 2014.

[32] 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.

[33] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. A large-scale car dataset for fine-grained categorization and verification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[34] Ning Zhang, Manohar Paluri, Yaniv Taigman, Rob Fergus, and Lubomir Bourdev. Beyond frontal faces: Improving person recognition using multiple cues. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4804–4813, 2015.

[35] 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*, 2015.