© IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.

This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

Domain Adaptation for CNN Based Iris Segmentation

1st Ehsaneddin Jalilian
Department of Computer Science
University of Salzburg
Salzburg, Austria
ehsaneddin.jalilian@sbg.ac.at

2nd Andreas Uhl
Department of Computer Science
University of Salzburg
Salzburg, Austria
andreas.uhl@sbg.ac.at

3rd Roland Kwitt
Department of Computer Science
University of Salzburg
Salzburg, Austria
roland.kwitt@sbg.ac.at

***Abstract*—Convolutional Neural Networks (CNNs) have shown great success in solving key artificial vision challenges such as image segmentation. Training these networks, however, normally requires plenty of labeled data, while data labeling is an expensive and time-consuming task, due to the significant human effort involved. In this paper we propose two pixel-level domain adaptation methods, introducing a training model for CNN based iris segmentation. Based on our experiments, the proposed methods can effectively transfer the domains of source databases to those of the targets, producing new adapted databases. The adapted databases then are used to train CNNs for segmentation of iris texture in the target databases, eliminating the need for the target labeled data. We also indicate that training a specific CNN for a new iris segmentation task, maintaining optimal segmentation scores, is possible using a very low number of training samples.**

***Index Terms*—Domain adaptation, CNN based iris segmentation, Iris segmentation**

I. Introduction

In recent years, considerable effort has been made towards developing accurate automatic segmentation systems for variety of applications, using supervised machine learning algorithms. Accurate segmentation of iris texture in eye images is a key challenge in iris recognition, and plays vital role in accuracy of subsequent feature extraction and recognition algorithms. Application of convolutional neural networks for iris segmentation has recently received some research attention, and a few CNN based models got proposed [1] [2]. Nonetheless, as any other supervised learning model, performance of these models are highly dependent on availability of sufficient amount of labeled data. Data labeling, however, is extremely expensive and time-consuming process, especially when segmenting iris data, due to the considerable human effort involved. As a result, manually annotating large number of data for each new segmentation task (i.e. new datasets or sensors, respectively) is not a feasible choice.

In this work, we propose two domain adaptation methods to transfer the domains of source iris databases (for which segmentation labels are available) to those of the targets, generating adapted iris databases, which in turn, enable training of a Fully Convolutional Neural Network (FCN) for segmentation of iris in the target databases. Doing so, we can train a FCN for a new iris segmentation task, using adapted source iris images and their corresponding ground-truth masks, eliminating the need for the target iris ground-truth masks, which are extremely expensive to generate. To address this objective, we selected three publicly available iris databases, and explored their tonal distribution in terms of the intensity values at pixel-level. Subsequently, we developed a linear and also a non-linear domain adaptation hypotheses to adapt the intensity information of source databases to those of the targets, generating a set of adapted databases. Eventually, we trained a FCN with the adapted databases, and then tested it on the target databases. At the end, we evaluated the expediency of our model by comparing the segmentation results obtained by the adaptation models against those of the cross- and within-databases.

978-1-4673-9563-2/15/$31.00 © 2015 IEEE

II. Related Works

Domain adaptation in computer vision is significantly focused on visual classification, with much research dedicated to generalizing across the domain transformation between images of objects and the same objects' photos in the real world [3] [4]. In this context, many of the researches concentrated on exploring feature representations which permute the greatest distractions between two domains [5] [6] [7]. Some other works tried to readjust such features by minimizing the distinction between their distributions [7] [8]. Liu et al. proposed a coupled generative adversarial network, to learn the joint distribution of images from both the source and the target databases [9].

Very limited research has been conducted on domain adaptation in other key computer vision fields such as detection and segmentation. To be more precise, in detection, Hoffman et al. introduced a domain adaptation model by explicitly modeling the representation shift between classification and detection models [10]. Also, in a follow-up work, they incorporated per-category adaptation using multiple instance learning [11]. The detection models were later converted into FCNs for evaluating semantic segmentation performance [12]. But this work did not propose any segmentation-specific adaptation approach. The only work with focus on CNN based segmentation is proposed by Hoffman et al. [13]. They used both source and target data in a fully-convolutional domain adversarial training, minimizing the global distance of feature space

between two domains. Then category updates were performed on the target images, using a constrained pixel-wise multiple instance learning objective. They used their model for semantic segmentation in city images obtained under different scenarios. The main drawback of their method is using adversarial training and shared weights. While applying this method lets the target network to adapt to the weights well, yet it degrades this process in the source network. As their experiments also show, while in most classes, they slightly improved the segmentation results, in some other classes such as "pole" and "truck" segmentation results show degradation.

III. Domain Adaptation for CNN Training

In this section, we describe our domain adaptation model for CNN based iris segmentation. Although without loss of generality, our approach is applicable to other segmentation models also. Given the source iris database X_s, and its corresponding ground-truths Y_s, $P(X_s)$ refers to the distribution of intensities in the source iris images. Likewise, we have the target iris images X_t, and their corresponding ground-truths Y_t, while $P(X_t)$ specifies the distribution of intensities in the target iris images. Under the domain difference scenario, we assume that the conditional distributions of Y_s and Y_t are the same, but the marginal distributions of X_s and X_t differ in the two domains. The resulting distinction between the distributions in two domains is refereed to as sample bias ϕ where

$$P_t = P_s(\phi(X_s), Y_s). \tag{1}$$

Using empirical risk minimization framework for supervised learning, we want to select an optimal parameter $\psi^{'}$, to minimize the following objective function

$$\psi_t^{'} = \arg\min_{\psi \in \Psi} \sum_{(x,y)\in X\times Y} \widetilde{P}_s(\phi(X_s), Y_s) g(x, y, \psi) \ , \tag{2}$$

$$\psi_t^{'} = \arg\min_{\psi \in \Psi} \sum_{i=1}^{N} g(\phi(x_s), y_s, \psi) \ , \tag{3}$$

where $g(x, y, \psi)$ is the loss function, and $\widetilde{P_s}(X, Y)$ is the empirical distribution of $P_s(X, Y)$. As it can be interpreted from (2), weighting the images' intensities of source data by ϕ provides the solution to the minimization function. The straight forward solution to weight the intensity values of source data is using a linear normalization model as follows:

$$b = (max(B) - min(B)) \frac{a - min(A)}{max(A) - min(A)} + (min(B)) \ , \tag{4}$$

where a and b are the input and output respectively, and $B = \{b_1, b_2, ..b_n\}$, and $A = \{a_1, a_2, ..a_n\}$. Our first (linear) domain adaptation method is based on the same model. In this way, we extracted the average range (maximum and minimum) of intensities in the iris, non-iris, and pupil regions of eye images in the target databases. Then using the above model (4), we weighted the intensity information of source databases to those of the targets, to generate new adapted databases as we already mentioned.

As it can be seen, this model provides a linear solution to our domain adaptation problem. In practice, in this method for each region, all the source intensity ranges get normalized to "a single average intensity range of that region in the target database." Yet it is a fact that, the intensity ranges of the target regions follow a non-linear distribution in the target databases. To address this non-linearity, we propose our second (non-linear) domain adaptation method. For this purpose, after extracting the maximums and minimums of each region in the target databases, we developed a probability distribution function (PDF) for each. To transfer the intensities in the source regions to those of the targets, initially we drew a random value from the corresponding PDFs, following a normal distribution.

However, this strategy seemed not to be so promising, as it neglected the complimentary relation between maximums and minimums in each region. Further analysis of the extracted intensities also revealed that there exists an obvious mutual relation between maximum and minimum intensity values in each region. So that, as the maximums increase, minimums also increase, and vise versa. To address this relation, after extracting the intensity ranges, for each unique maximum value, we calculated the mean of corresponding minimum values. Then we developed a cross-value matrix for each region, using these two variables. Next, we applied kernel smoothing regression to this data to generate a polynomial function $f(X)$ as follows:

$$f(x) = p_1 x^n + p_2 x^{n-1} + ... + p_n x + p_{n+1} \ , \tag{5}$$

where x represents the input (minimum) to the model, and n is the degree of polynomial function. Now, to adapt the domain of each source image to that of the targets, we randomly selected a minimum for each region, and then estimated its corresponding maximum using the polynomial model we proposed, as demonstrated in figure 1. As a result, unlike in the linear adaptation method, where all images were mapped to the same range, here each adapted image has a potentially different range.

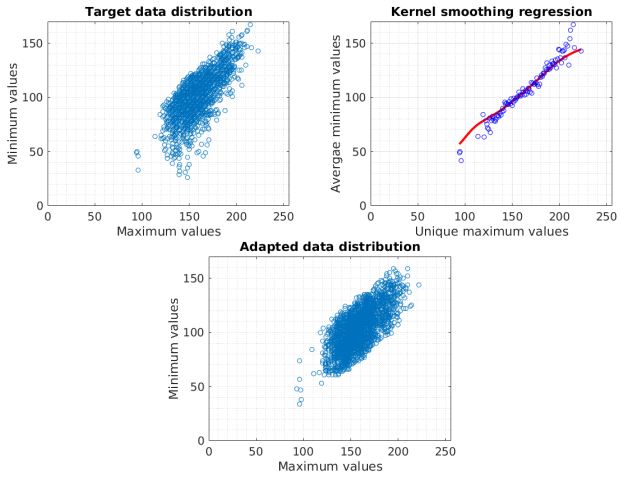

Fig. 1. Sample non-linear data adaptation steps

IV. Experimental Framework

To assess the expediency of our domain adaptation methods we carried out a set of segmentation experiments on the databases. The details of these experiments are explained in the next section respectively. Yet, in the following we explain the framework for these experiments.

A. Databases

For our segmentation experiments we used three publicly available iris databases. The Casia-iris-interval-v4 (Casia4i) database [1], which contains 2640 iris images belonging to 249 subjects. The iris images in this database were acquired under near-infrared illumination. The IITD database (Iitd) [2], which consists 2240 iris images corresponding to 224 subjects. All these images are acquired in indoor environment, in near-infrared illumination. The Casia-iris-aging-v5 database, which is a subset of the upcoming Casia-v5 (Casia5a) iris database [3], contains 120 images per eye and user from video sequences captured in 2009, and 20 images per eye and user from video sequences captured in 2013. The segmentation ground-truth masks for these databases were provided by the University of Salzburg [4].

B. FCN Netwrok

The architecture of the network we used in this work is similar to the basic fully convolutional encoder-decoder network proposed by Kendall et al. [14]. However, we re-designed the softmax layer to segment the iris and non-iris areas only. The network's encoder architecture is organized in four stocks, containing a set of blocks. Each block comprises a convolutional layer, a batch normalization layer, and a rectified-linear non-linearity layer. The corresponding decoder architecture, likewise, is organized in four stocks of blocks, whose layers are similar to those of the encoder blocks, except that here each block includes an up-sampling layer also. The decoder network ends up to a softmax layer which generates the final segmentation map. More details about the technical specification of the network and layers can be found in the relevant reference. The network was implemented in the "Caffe" deep learning framework.

C. Metrics and Measurements

We estimated iris segmentation accuracies using two segmentation error scores of nice1 $(n1)$ and nice2 $(n2)$, which are based on the NICE.I protocol [5]. The error score nice1 calculates the proportion of corresponding disagreeing pixels (by the logical exclusive-or operator) over all the image as follows:

[1]Chinese Academy of Sciences, Institute of Automation, Center for biometrics and security research, http://biometrics.idealtest.org

[2]Indian Institute of Technology Delhi, IIT Delhi Iris Database, http://www4.comp.polyu.edu.hk/ csajaykr/database.php

[3]see http://www.biometrics.idealtest.org

[4]http://www.wavelab.at/sources/Hofbauer14b

[5]http://nice1.di.ubi.pt/dates.htm

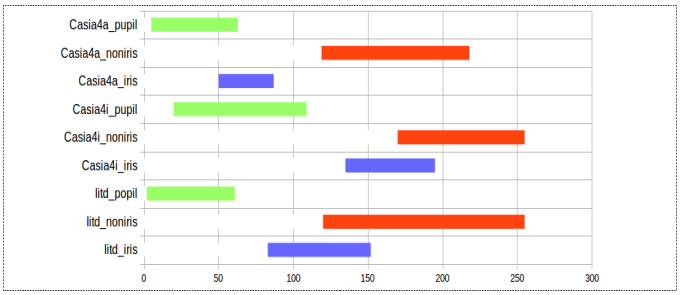

Fig. 2. Average intensity ranges of iris, non-iris, and pupil regions in databases

$$nice1 = \frac{1}{c \times r} \sum_{c'} \sum_{r'} O(c', r') \otimes C(c', r') \ , \qquad (6)$$

where c and r are the columns and rows of the segmentation masks, and $O(c', r')$ and $C(c', r')$ are, respectively, pixels of the output and the ground-truth mask. The error score nice2 intends to compensate the disproportion between the priori probabilities of iris and non-iris pixels in the images - it averages type-I and type-II errors, i.e. between the fp (false positives) and fn (false negatives) rates as follows:

$$nice2 = \frac{1}{2} \ (fp + fn) \ . \qquad (7)$$

Additionally, we considered the F score $(f1)$ to estimate iris segmentation accuracies also. The F score is the harmonic mean of precision (P) (the fraction of relevant instances among the retrieved instances) and recall (R) (the fraction of relevant instances that have been retrieved over total relevant instances) as follows:

$$f1 = 2 \ \frac{RP}{R + P} \ . \qquad (8)$$

The values of nice1 and nice2 are bounded in [0, 1] interval, and "1" and "0" are respectively the worst and the best scores. The F score values are bounded in [1, 0] interval, and "0" and "1" are the worst, and the best scores respectively.

V. Experiments and Discussions

We evaluated the eminence of our domain adaptation model by a set of experiments. In this way, initially we developed six sets of unique database pairs (source-target), using three available databases. Next, we explored the distributions of domains in the target databases, extracting the intensity ranges of iris, non-iris, and pupil regions of eye images in these databases (figure 2 reflects these information). Then, using our domain adaptation methods, we transfered the intensity values of the specified regions in source databases to those of the targets, to produce an adapted database for each pair. Next, we trained our network with each adapted database, and then tested it on the corresponding target databases (adapted-target). Figure 3 and figure 4 show sample adapted images and their corresponding segmentations results for three database

TABLE I
SEGMENTATION SCORES FOR THE LINEAR-BASED (LB), AND NON-LINEAR-BASED (NB) DOMAIN ADAPTATION METHODS AGAINST THE BASELINE (SOURCE-TARGET) RESULTS

Method	Adapted-target (LB)			Adapted-target (NB)			Baseline(Source-target)		
Scores	nice1	nice2	f1	nice1	nice2	f1	nice1	nice2	f1
Casia5a-casia4i	0.186	0.220	0.610	0.274	0.353	0.098	0.292	0.640	0.003
Casia5a-iitd	0.148	0.172	0.781	0.266	0.305	0.498	0.229	0.221	0.473
Casia4i-casia5a	0.066	0.194	0.730	0.027	0.074	0.859	0.274	0.406	0.341
Casia4i-iitd	0.121	0.141	0.808	0.102	0.095	0.812	0.218	0.219	0.724
Iitd-casia5a	0.062	0.185	0.739	0.034	0.088	0.813	0.049	0.117	0.830
Iitd-casia4i	0.299	0.319	0.569	0.208	0.174	0.374	0.315	0.584	0.045

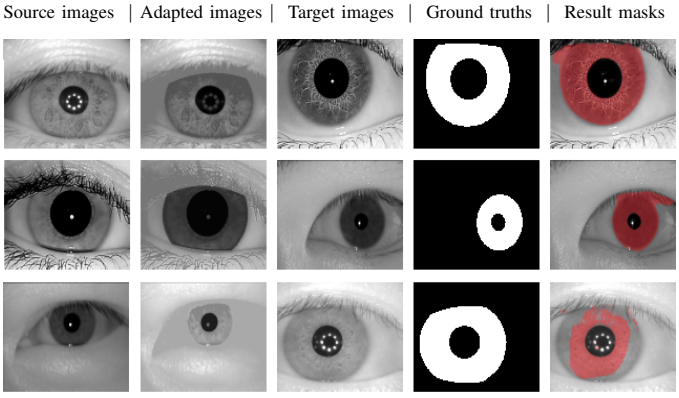

Fig. 3. Sample adapted images and their corresponding segmentation results for Casia4i-Iitd (first row), Iitd-Casia5a (second row), and Casia5a-Casia4i (third row) database pairs (source-target) using the linear domain adaptation method

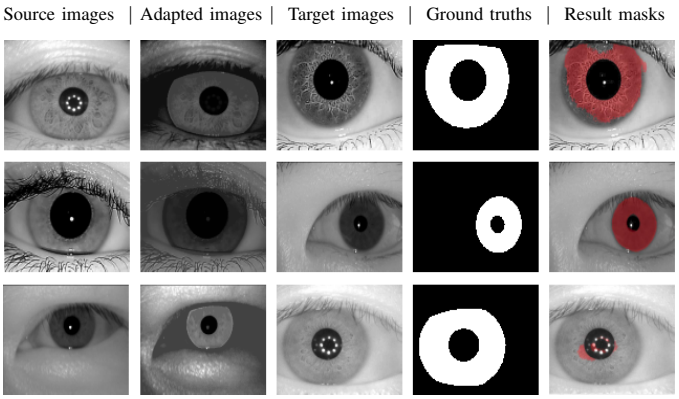

Fig. 4. Sample adapted images and their corresponding segmentation results for Casia4i-Iitd (first row), Iitd-Casia5a (second row), and Casia5a-Casia4i (third row) database pairs (source-target) using the non-linear domain adaptation method

pairs, applying linear and non-linear adaptation methods respectively.

The segmentation results then were compared against the baseline results (results of applying network trained with the source databases directly to the target databases without adaptation). Table 1 shows the segmentation scores for our linear-based (LB), and non-linear-based (NB) adaptation methods against the baseline (source-target) results. In addition, figure 5 provides further information including: min, max, median, quantiles, and outliers for the liner-based adaptation experiments in form of box-plots.

As the experiment results in Table 1 show, almost all linear domain adaptations result in significant improvement of iris segmentations compared to the baseline results. Slightly lower, yet stable improvements can also be seen in the segmentation results of non-linear domain adaptations. It should be noted that feature representations affecting the weights during training process are not limited to tonal distributions, and further features such as geometric properties of iris, non-iris, and pupil regions are definitely affecting this process. Here we just considered the tonal distributions, so the results are not comparable with the optimal solution when directly training with the target dataset.

All in all, the overall results confirm the key conclusion that tonal distribution (intensity ranges of iris, non-iris, and pupil) plays a key role in generalization of FCNs on new iris data that differs from the training data. It is also interesting to note that, while the segmentation results for linearly adapted Iitd-casia5a databases show slightly lower scores than the baseline, yet the segmentation results for non-linear adaptation of the same databases score much better compared to those of the baseline. Similar affinity can be found in the segmentation results of Casia5a-iitd databases, but in reverse manner.

While the proposed domain adaptation methods proved to effectively transfer the domains between the iris databases, yet the segmentation results obtained are far from the optimal iris segmentation scores as demonstrated in Table 2. To this extent, with the aim of minimizing the number of labeled data required to train the CNNs for new iris segmentation tasks, and maintaining optimal segmentation scores, we conducted a series of additional experiments. In this way, we decreased the number of labeled samples required to train a CNN for a new iris segmentation task stepwise, obeying the framework we used for our optimal (target-target) experiments. Table 3

TABLE II
OPTIMAL (TARGET-TARGET) SEGMENTATION RESULTS

Method	Target-Target		
Scores	nice1	nice2	f1
Casia5a-Casia5a	0.019	0.038	0.925
Casia4i-Casia4i	0.033	0.038	0.937
Iitd-Iitd	0.027	0.032	0.951

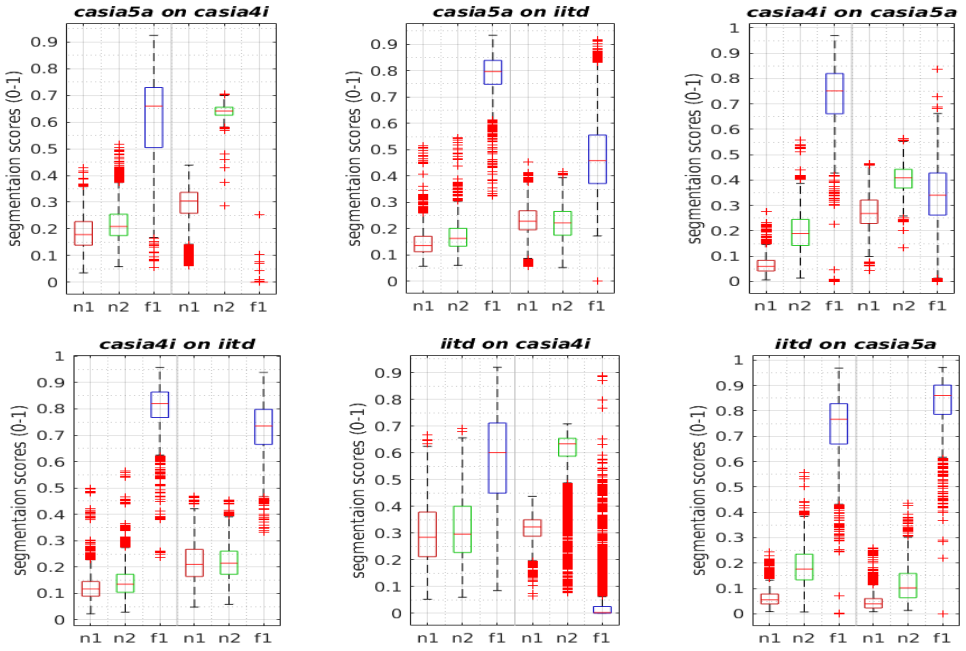

Fig. 5. Segmentation results for the linear domain adaptations (left side of graphs), against the baseline results (right side of graphs)

TABLE III
SEGMENTATION RESULTS FOR DECREASED NUMBER OF TRAINING SAMPLES

Database	Casia5a			Casia4i			Iiitd		
Score	nice1	nice2	f1	nice1	nice2	f1	nice1	nice2	f1
15 pcs	0.075	0.082	0.875	0.205	0.263	0.502	0.089	0.097	0.856
25 pcs	0.064	0.077	0.896	0.099	0.115	0.814	0.077	0.083	0.879
50 pcs	0.050	0.070	0.909	0.078	0.068	0.841	0.063	0.070	0.889
100 pcs	0.021	0.040	0.921	0.038	0.039	0.926	0.035	0.037	0.941

demonstrates the results for these experiments. Considering the optimal segmentation results in Table 2, we can see that for most databases optimal segmentation scores can be achieved using maximum number of 100 training samples. However, in most cases slightly lower, but very close scores can be achieved with 50 or even 25 samples.

VI. CONCLUSION

Application of convolutional neural networks for iris segmentation has recently received first research attention, and some CNN based models got introduced for this purpose by researchers. Nonetheless, as any other supervised learning model, training these models require adequate amount of labeled iris data. Due to the significant human effort involved, preparing labeled data to train these networks for new segmentation tasks is very expensive and time consuming. In this work, we proposed two adaptation methods to transfer the domains of source iris databases to those of the targets, producing adapted databases. The adapted iris images along with their corresponding ground-truth masks then enabled training of a FCN network for segmentation in target iris databases, eliminating the need for the target ground-truth masks.

While experimental results proved expediency of these two methods, yet in some cases, their segmentation scores were far from the optimals. With the aim of minimizing the number of labeled iris images required to train the network for new iris segmentation tasks, and also maintaining optimal segmentation scores, we decreased the number of training samples stepwise as an alternative approach to domain adaptation. The experiments demonstrated that for most databases, optimal segmentation scores can be achieved using maximum of 100 training data samples. In our future work, we will investigate the relations between the two proposed adaptation methods and the reasons for the different results. Beside this, we will explore more feature representations which encourage maximal distinction between two domains, hoping to be able to develop a more comprehensive domain adaptation method.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 700259.

REFERENCES

[1] E. Jalilian and A. Uhl, "Iris segmentation using fully convolutional encoder–decoder networks," in *Deep Learning for Biometrics*. Springer, 2017, pp. 133–155.

[2] N. Liu, H. Li, M. Zhang, J. Liu, Z. Sun, and T. Tan, "Accurate iris segmentation in non-cooperative environments using fully convolutional networks," in *Biometrics (ICB), 2016 International Conference on*. IEEE, 2016, pp. 1–8.

[3] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," *Computer Vision–ECCV 2010*, pp. 213–226, 2010.

[4] B. Kulis, K. Saenko, and T. Darrell, "What you saw is not what you get: Domain adaptation using asymmetric kernel transforms," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1785–1792.

[5] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 4068–4076.

[6] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *Journal of Machine Learning Research*, vol. 17, no. 59, pp. 1–35, 2016.

[7] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *International Conference on Machine Learning*, 2015, pp. 97–105.

[8] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 136–144.

[9] M.-Y. Liu and O. Tuzel, "Coupled generative adversarial networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 469–477.

[10] J. Hoffman, S. Guadarrama, E. S. Tzeng, R. Hu, J. Donahue, R. Girshick, T. Darrell, and K. Saenko, "Lsda: Large scale detection through adaptation," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3536–3544. [Online]. Available: http://papers.nips.cc/paper/5418-lsda-large-scale-detection-through-adaptation.pdf

[11] J. Hoffman, D. Pathak, T. Darrell, and K. Saenko, "Detector discovery in the wild: Joint multiple instance and representation learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2883–2891.

[12] J. Hoffman, D. Pathak, E. Tzeng, J. Long, S. Guadarrama, T. Darrell, and K. Saenko, "Large scale visual recognition through adaptation using joint representation and multiple instance learning," *Journal of Machine Learning Research*, vol. 17, no. 142, pp. 1–31, 2016.

[13] J. Hoffman, D. Wang, F. Yu, and T. Darrell, "Fcns in the wild: Pixel-level adversarial and constraint-based adaptation," *arXiv preprint arXiv:1612.02649*, 2016.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *arXiv preprint arXiv:1511.00561*, 2015.