Cross-domain Visual Attention Model Adaption with One-shot GAN

Daowei Li[1] Kui Fu[1] Yifan Zhao[1] Long Xu[2] Jia Li[1*]
[1]State Key Laboratory of Virtual Reality Technology and Systems, SCSE, Beihang University
[2]Key Laboratory of Solar Activity, National Astronomical Observatories, CAS, Beijing, China
[1]{lidaowei, kuifu, zhaoyf, jiali}@buaa.edu.cn [2]lxu@nao.cas.cn

Abstract

The state-of-the-art models for visual attention prediction perform well in common images. But in general, these models have a performance degradation when applied to another domain with conspicuous data distribution differences, such as solar images in this work. To address this issue and adopt these models from the common images to the sun, this paper proposes a new dataset, named VASUN, that records the free-viewing human attention on solar images. Based on this dataset, we propose a new cross-domain model adaption approach, which is a siamese feature extraction network with two discriminators and trained in a one-shot learning manner, to bridge the gaps between the source domain and target domain through the joint distribution space. Finally, we benchmark existing models as well as our work on VASUN and give some analysis about predicting visual attention on the sun. The results show that our method achieves state-of-the-art performance with only one labeled image in the target domain and contributes to the domain adaption task.

1. Introduction

Visual attention prediction, which simulated the human vision system to quickly pay attention to parts of the image instead of the whole scene pin its entirety, has received increasing attention in the past two decades, *e.g.*, bio-inspired attention models [15, 12, 35], shallow learning models [18, 19], and deep learning models [20, 14, 26, 28, 32]. These models have been shown to perform very well when tested on the common data related to the training data (what we call the source domain), but their performance drops dramatically when applied them to abnormal data (*e.g.*, solar images). The challenge is that the distribution of the source domain and the target domain are very different. We will approach this challenge by investigating how a common representation between the source domain and the target domain can map the two domains to have similar distributions, enabling effective domain adaption.

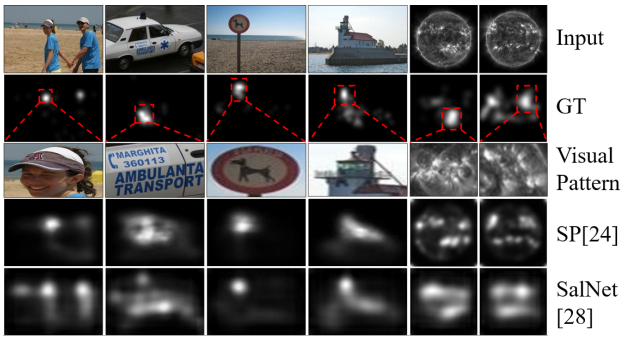

Figure 1. Representative results of visual attention models on common images and solar images.

Although existing state-of-the-art models have effective domain adaption among these common image sence, it is questionable whether these models have an aptitude for memorizing certain features and intermediate representations related to common images, or they have strong generalization ability. Intuitively, as shown in Fig. 1, these models perform well on conventional datasets but not on a solar dataset. In order to investigate how a common representation between the source domain and the target domain can make the two domains appear to have similar distributions, we firstly propose a new visual attention dataset, the images of which are provided by the LSDO dataset [21]. Based on this dataset, we propose weakly supervised domain adaption with one-shot GAN to address this problem.

We also benchmark the performance of 16 visual attention models as well as our method over the new dataset by using five representative evaluation metrics. The benchmark results show that our model achieves state-of-the-art performance only by introducing one labeled image in target domain to the network.

Our contributions are summarized as follows: 1) We propose novel visual attention dataset covering 1070 solar images. To the best of our knowledge, the proposed dataset is

*Jia Li is the corresponding author.

Table 1. The subject number and image resolution of representative image datasets and our VASUN dataset. #Img and #Sub mean the number of images and subjects in the dataset.

Dataset	#Img	#Sub.	Max Res.
MIT300 [17]	300	39	1024×1024
MIT1003 [18]	1003	15	1024×1024
Toronto [3]	120	20	511×681
CAT2000 [2]	2000	18	1920×1080
SALICON [16]	15000	-	640×480
PASCAL-S [25]	850	8	500×500
DUT-O [33]	5168	5	401×401
VASUN	1070	16	1024×1024

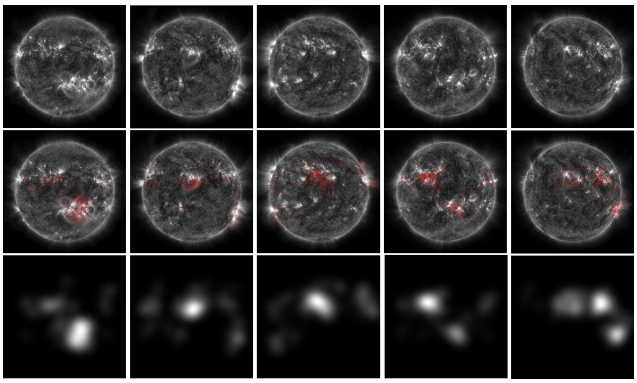

Figure 2. Representative examples from VASUN. Red dots in the second row indicate the recorded fixation points, which are used to generate the ground-truth attention maps in the third row.

the first eye-tracking dataset for the solar attention prediction.2) We propose a novel cross-domain model adaption network, which consists of a siamese feature extraction network and two generative adversarial networks and trained in a one-shot learning manner. 3) We present a comprehensive analysis of an extensive benchmark and the results give a positive evaluation about our cross-domain visual attention model adaption.

2 The VASUN Dataset

Many attention benchmark datasets have been proposed in the literature. Among these datasets, some of them are specifically designed for studying the human fixation prediction problem of more than a dozen subjects in the free-viewing conditions (*e.g.*, MIT300 [17], MIT1003 [18], Toronto [3] and CAT2000 [2]), while the rest ones just record fixations of several subjects to facilitate subsequent annotations (*e.g.*, Pascal-S [25], and DUT-O [33]). The number of subjects and their image resolutions of these

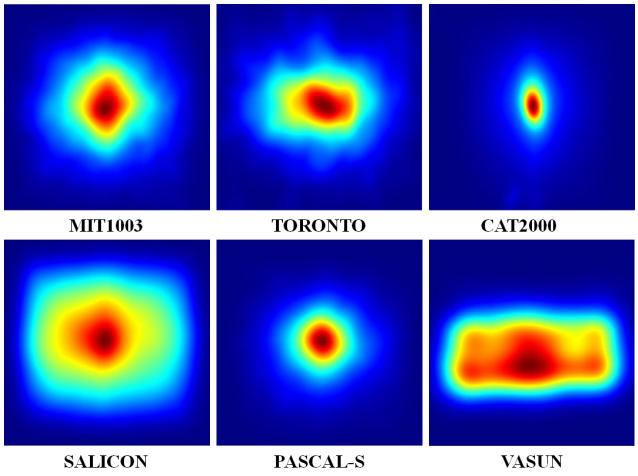

Figure 3. The average attention maps of various datasets.

datasets can be found in Tab. 1.

To measure the generalization ability and to make the attention prediction results in more valued for application, we construct a dataset that contains new visual patterns in solar images. Toward this end, we refer to the solar images in LSDO dataset [21]. Neither such solar images nor their visual patterns have appeared in any existing image attention datasets. From the LSDO dataset, we construct a new image attention dataset to study visual attention on the sun (denoted as VASUN). We adopt a similar setting with MIT1003, the most widely used dataset with daily scenarios. That is, we sample 1070 images taken at the wavelength 171Å from LSDO and down-sample them to the resolution of 1024×1024. On these images, we record the visual attention of 16 subjects in eye-tracking experiments. These 16 subjects (10 males and 6 females) have normal or correct-to-normal visions. None of them has prior knowledge of astronomy and solar physics to avoid subjective bias.

Some representative examples of the recorded fixations and ground-truth attention maps can be found in Fig. 2. From this figure, we can see that most human visual attention is allocated to large active regions, making the visual attention prediction task on solar images a theoretically simple task. In addition, we also compare the average attention maps of VASUN with previous datasets in Fig. 3. We can see that the distribution of fixations in VASUN is quite different from previous datasets with daily scenarios. This may be caused by the solar images have no photographer bias that tend to place the target at image centers. This phenomena further validates that VASUN can be used to test the generalization ability of visual attention models.

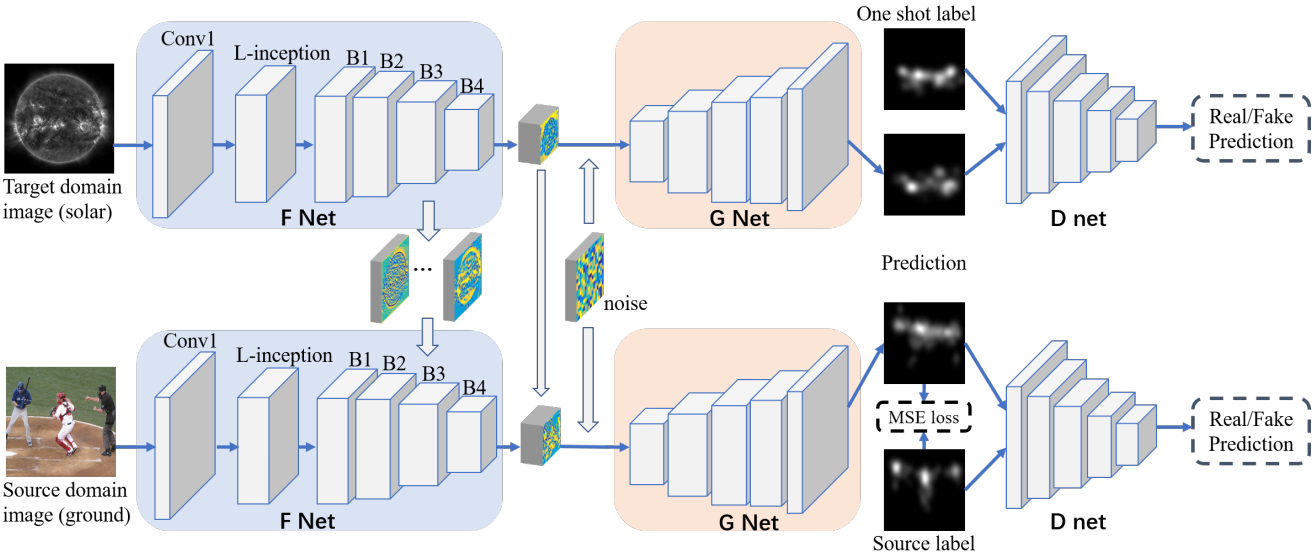

Figure 4. The network architectures of our proposed model. Given a pair of images (x_i, x_0), target stream F network extracts features f_{x_0} and transforms them to source F network hierarchically. Generator G maps the embedding x_{0g}, x_{ig} to the attention maps for target image and source image respectively. Then D network discriminates the maps real or not. MSE supervisor is applied to make sure the correctness of prediction in source path. Note that the F, G, and D share weights in the siamese paths.

3 Model Adaption with One-shot GAN

3.1 Problem Description

Before discussing the details of our proposed method, we firstly describe the problem to solve. Let $\mathbf{X}_s = \{x_i\}_{i=1}^N$, $\mathbf{Y}_s = \{y_i\}_{i=1}^N$ be the input images and their corresponding labels in source domain space. While the $\mathbf{X}_t = \{x_i\}_{i=1}^M$, $\mathbf{Y}_t = \{y_i\}_{i=1}^M$ represent the input images and labels in target domain space. $\mathbf{Y} = \mathbf{Y}_s \cup \mathbf{Y}_t = \{y_i\}_{i=1}^{N+M}$ refers to the joint labels of source and target domain. The best adapted model $H(x)$ can be formulated as

$$H(x) = \lambda \cdot F_s(x) + (1-\lambda) \cdot F_t(x), \lambda = \begin{cases} 1, x \in X_s, \\ 0, x \in X_t, \end{cases} \quad (1)$$

where F_s and F_t represent the mapping relation between input images and their labels in the source domain and target domain, respectively. The parameter λ is used to balance F_s and F_t. In the one-shot learning for domain adaption manner, since we only have one labeled data from $\mathbf{X}_t$, denoted as (x_0, y_0), it is difficult to learn the function F_t for the target domain. However, the mapping F_s can access the source distribution using labeled data from $\mathbf{X}_s$. As a result, the problem can be stated as learning a predictor that is optimal in the joint distribution space by using labeled source data and one sample labeled data from the target domain. We change the data organization mode from separated $\mathbf{X}_s$, $\mathbf{X}_t$ into $\mathbf{X} = \{(x_i, x_0)\}_{i=1}^N$, where the (x_i, x_0) means a pair of images. Let $\mathbf{R}_d$ be the common representation. Then our goal is to learn an embedding map $F : \mathbf{X} \mapsto \mathbf{R}_d$ and a

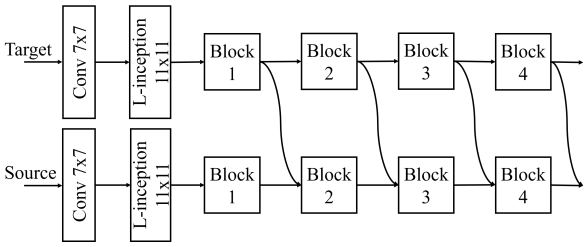

Figure 5. The network architectures of siamese F network.

prediction function $G : \mathbf{R}_d \mapsto \mathbf{Y}$. Both F and G are deep neural networks. As a result, during the training process, F use the features extracted from the source domain and the representative image of target domain to predict attention maps on the source domain, which helps F reduce the domain shift between the distributions of source and target domain. The (1) can be modified as

$$H(x) = G(\boldsymbol{\lambda} \cdot F(x_i) + (1-\boldsymbol{\lambda}) \cdot F(x_0)), x = (x_i, x_0), \quad (2)$$

where balance parameter $\boldsymbol{\lambda}$ is learnable.

Several Previous works have provided ways to transfer information between the source and target distributions, including learning entropy-based metrics [27], learning a domain classifier based on an embedding network [8] or denoising autoencoders [10]. In this work, we propose a GAN-based approach to bridge the gaps between the source domain and the target domain.

Algorithm 1 Iterative training procedure of our approach

training iterations = N
for each $i \in [1, N]$ **do**
  Given a pair of images (x_i, x_0) and labels(y_i, y_0);
  Let $f_i = F(x_i)$ be the source embedding;
  Let $f_0 = F(x_0)$ be the target embedding;
  Sample k random noise samples$\{z_i\}_{i=1}^{k} \in N(0, 1)$;
  Let x_{ig}, x_{0g} be the concatenated input to generator;
  Update discriminator D with following objectives:

$$\begin{aligned} L_D = &\max_D \frac{1}{k}\sum_{i=1}^{k} \log^{D(x_i)} + \log^{1-D(G(x_{ig}))} \\ &+\frac{1}{k}\sum_{i=1}^{k} \log^{D(x_0)} + \log^{1-D(G(x_{0g}))} \end{aligned} \tag{3}$$

  Update generator G, loss comes from strong supervisor and discriminator:

$$\begin{aligned} L_G = &\min_G \frac{1}{k}\sum_{i=1}^{k} \|G(x_{ig}) - y_i\|_2^2 \\ &+\frac{1}{k}\sum_{i=1}^{k} \log^{1-D(G(x_{ig}))} + \log^{1-D(G(x_{0g}))} \end{aligned} \tag{4}$$

  Updating embedding network F using strong supervisor loss and a linear combination of the adversarial loss:

$$\begin{aligned} L_F = &\min_F \frac{1}{k}\sum_{i=1}^{k} \|G(x_{ig}) - y_i\|_2^2 \\ &+\frac{1}{k}\sum_{i=1}^{k} \alpha \log^{1-D(G(x_{ig}))} \\ &+\beta \log^{1-D(G(x_{0g}))} \end{aligned} \tag{5}$$

end for

3.2 Proposed Approach

In this work, we propose a Cross-domain Model Adaptipon with one-shot GAN for Attention prediction, denoted as CMA-GAN, which utilizes a variant of the typical GAN to adapt the attention prediction model from daily scenarios to solar image domain. The overall structure of our network is given by Fig. 4 and detailed siamese F network is illustrated in Fig. 5. We will describe the model in details next:

The network F directly extract features from the input image pairs (x_s, x_0), and then the input of the generator network G can be represented as $x_g = [\boldsymbol{\lambda} \cdot F(x_s) + (1 - \boldsymbol{\lambda}) \cdot F(x_0), z]$, which is a combination of the features extracted from the source domain input x_i, target domain input x_0 and a random noise vector $z \in \mathbf{R}_d$ sampled from $N(0, 1)$.

We employ a decoder network G as the generator which takes the embedding generated from F as input and produce the attention map as its output $G(x)$. Then $G(x)$ is delivered to the discriminator to judge the probability that x belongs to its domain distribution. In order to make the F and G correctly predict the attention on daily scenario images, we also use the strong supervised manner (*MSE loss* in this paper) with labels in $\mathbf{Y}_s$.

We take coupled discriminators to judge the probability that the input image pair(x_s, x_0) are real or not, which referred as $D(x_s)$ and $D(x_0)$, respectively. In order to avoid *over-fitting*, these two discriminators share wights in training. Since we have known the label for x_0, $D(x_s)$ and $D(x_0)$ are both used to back-propagate the gradients. However they have different weights in different stage. To jointly learn the embedding and the generator-discriminator pair, we optimize the G, F, G as the Alg. 1:

From Alg. 1, we can see that the discriminator network are optimized by minimizing two cross binary loss $L_{ds,s}$ and $L_{dt,t}$ which come from $D(x_s)$ and $D(x_0)$, respectively. Then the generator networks are optimized with adversarial loss and the strong supervisor *MSE* loss. Feature extraction networks are updated in the same way as the generator. However, we set different weights for two adversarial losses with $\alpha = 0.2$, $\beta = 0.8$ to give the embedding extracted from the target images higher priority to make sure a better concatenation between thep source and target distribution. As for Fig. 5, we show the transformation of features between siamese F network in detail. In addition, we take all the key factors, which are big kernels in low layers, appropriate depth of network and multi-scale input (three scales of the image in our model) analyzed in the previous experiment into consideration.

4 Experiment

4.1 Training and Testing

Our experiments are conducted with the *pytorch* toolbox. During the training phase, the learning rate is set to 10^{-3}. To verify the effectiveness of our approach, we design another three control models for ablation study, which are *Resnet50-backbone with source only*, *Resnet50-backbone with one-shot learning*, *L-inception Resnet50-backbone with source only*, respectively. We organized study by carrying out these experiments as follow phases:

We reorganize the dataset SALICON with our VASUN to train these control models from scratch. For the training of source only model, we keep the original structure of SALICON which has 10,000 training images. While for the training of the one-shot model, we first select one labeled solar image. To improve the performance of proposed approach, We reconciled the following two constraints: the number of events on the face of the selected image should

Table 2. Performance of models in different phase on VASUN-testing. The best models of each column are marked with bold.

Model	AUC	sAUC	NSS	SIM	CC
res50-source	0.856	0.717	1.263	0.486	0.540
res50-oneshot	0.896	0.741	1.451	0.570	0.622
lres50-source	0.879	0.736	1.311	0.537	0.562
CMA-GAN	**0.899**	**0.775**	**1.585**	**0.598**	**0.675**
human	1.000	0.900	2.347	0.988	1.000

be as much as possible; the distance between ground truth map of the selected image and other maps should be short as much as possible. Then we organize the SALICON and the selected image as one-shot training dataset.

Then we train these control models and our network on the same platform in order. For the source only models, which are a typical encoder-decoder network, we train them on the SALICON training dataset end verify them with several solar images to gain their best performance. While for the one-shot models, including our approach, we train them on the one-shot training dataset and verify them on the same solar images to make sure they have learned the features in sun and are in the best condition.

After training, all models are changed into the testing structure and initialized with parameters obtained in training phase. The we evaluate them on the same testing set of VASUN. The performance is presented in Tab. 2.

From Tab. 2, we can see that *res50-source* has the worst performance which can be attributed to its simple architecture and single training manner. In contrast, the *res50-oneshot* has obvious performance improvements, increasing from 1.263 to 1.451, which results in a gain of 14.89% in NSS score. We can safely conclude that our proposed idea does work on the task of cross-domain attention model adaption. While the comparison between *res50-source* and *lres50-source* shows that the key factors we utilize to enhance the feature extract network also contribute to the improvement of the model's performance. Finally, we can see that our approach which integrates all the advantages discussed in this paper result in the best performance of visual attention prediction on the sun. To further evaluate our work's performance, we generate a new attention model benchmark based on our VASUN.

4.2 Model Benchmark

Many typical works have been proposed in the last few decades since attention prediction is a classic computer vision task. In this section, we refer to many famous or latest method and benchmark them on the VASUN to measure their generalization ability and evaluate the performance of our approach.

Table 3. Benchmark of 17 models with default parameters. The best models of each column are marked with bold, the best models of each group and each column are marked with a underline.

	Models	AUC	sAUC	NSS	SIM	CC
BIO	SUN [35]	0.844	0.664	1.135	0.443	0.488
	BMS [34]	0.838	0.717	1.333	0.500	0.569
	COV [5]	0.792	0.679	1.139	0.378	0.487
	GBVS [12]	0.821	0.673	1.146	0.431	0.489
	CAS [11]	<u>0.893</u>	0.722	1.372	0.515	0.584
	AWS [9]	0.871	0.743	1.474	<u>0.537</u>	0.629
	HFT [23]	0.887	<u>0.749</u>	<u>1.528</u>	0.522	<u>0.653</u>
SL	ICL [13]	<u>0.874</u>	<u>0.767</u>	<u>1.564</u>	0.556	<u>0.665</u>
	SSD [22]	0.787	0.651	1.342	0.398	0.427
	LDS [7]	0.860	0.734	1.553	<u>0.571</u>	0.656
	FES [30]	0.759	0.658	1.066	0.447	0.448
DL	eDN [31]	0.837	0.698	1.078	0.378	0.464
	iSEEL [29]	0.841	0.648	0.987	0.451	0.425
	SalNet [28]	0.877	0.729	1.361	0.512	0.580
	SALICON [14]	0.857	0.705	1.400	<u>0.537</u>	0.592
	SAM-ResNet [4]	<u>0.894</u>	<u>0.743</u>	<u>1.408</u>	0.533	<u>0.599</u>
	CMA-GAN	**0.899**	**0.775**	**1.585**	**0.598**	**0.675**

Over all the attention models, we benchmark 16 visual attention models from them, which can be roughly divided into three groups: 1) the **BIO** group contains seven bio-inspired models, including SUN [35], BMS [34], COV [5], GBVS [12], CAS [11], HFT [23] and AWS [9]; 2) the **SL** group contains five shallow learning models, including ICL [13], SP [24], SSD [22], LDS [7] and FES [30]; 3) the **DL** group contains five deep learning models, including eDN [31], iSEEL [29], SalNet [28], SALICON [14] and SAM-ResNet [4]. All these models have public source code on the Internet, and we use their default parameters to generate the attention maps. The predictions of these models are evaluated using five metrics, including AUC, sAUC, NSS, SIM and CC. The performance is presented in Tab. 3.

From Tab. 3, we find that in previous works, deep attention models on daily scenario images outperform shallow learning models and bio-inspired models. The key issue here is that the hand-craft features designed for daily scenarios may be not suitable for solar images and the deep attention models still cannot represent the target domain knowledge without re-training or fine-tuning. However, after introducing our idea which uses the generative adversarial network with one-shot learning manner, the performance of the attention model trained mainly by the daily scenario images shows obvious improvement in terms of all evaluation metrics.

5 Conclusions

In this paper, we revisit the problem of visual attention prediction from a novel perspective: the adaption of existing deep attention models. We propose a new dataset of solar images and a new approach to achieve the cross-domain attention model adaption with one-shot GAN under the circumstances that the target domain cannot provide enough labeled images to re-train the network. In addition, we hope our work can arise other researchers' interests in solar images and provide a feasible approach for cross-domain attention model adaption.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China under Grants 61672072 and 11790305, and also supported by the Beijing Nova Program (Z181100006218063).

References

[1] I. M. Author. Some related article I wrote. *Some Fine Journal*, 99(7):1–100, January 1999.

[2] A. Borji and L. Itti. Cat2000: A large scale fixation dataset for boosting saliency research. *CVPR 2015 workshop on "Future of Datasets"*, 2015.

[3] N. D. B. Bruce and J. K. Tsotsos. Saliency based on information maximization. In *NIPS*, pages 155–162, 2005.

[4] M. Cornia, L. Baraldi, G. Serra, and R. Cucchiara. Predicting human eye fixations via an lstm-based saliency attentive model. *IEEE TIP*, 2018.

[5] E. Erdem and A. Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. *Journal of vision*, 13(4):11–11, 2013.

[6] A. N. Expert. *A Book He Wrote*. His Publisher, Erewhon, NC, 1999.

[7] S. Fang, J. Li, Y. Tian, T. Huang, and X. Chen. Learning discriminative subspaces on random contrasts for image saliency analysis. *IEEE TNNLS*, 28(5):1095–1108, 2017.

[8] Y. Ganin and V. Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.

[9] A. Garciadiaz, V. Leborn, X. R. Fdezvidal, and X. M. Pardo. On the relationship between optical variability, visual saliency, and eye fixations: a computational approach. *Journal of Vision*, 12(7):17, 2012.

[10] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li. Deep reconstruction-classification networks for unsupervised domain adaptation. In *ECCV*, pages 597–613. Springer, 2016.

[11] S. Goferman, L. Zelnik-Manor, and A. Tal. Context-aware saliency detection. *IEEE TPAMI*, 34(10):1915–1926, 2012.

[12] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. In *NIPS*, 2007.

[13] X. Hou and L. Zhang. Dynamic visual attention: Searching for coding length increments. In *NIPS*, 2009.

[14] X. Huang, C. Shen, X. Boix, and Q. Zhao. Salicon: Reducing the semantic gap in saliency prediction by adapting deep neural networks. In *IEEE ICCV*, 2015.

[15] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE TPAMI*, 20(11):1254–1259, 1998.

[16] M. Jiang, S. Huang, J. Duan, and Q. Zhao. Salicon: Saliency in context. In *CVPR*, pages 1072–1080, 2015.

[17] T. Judd, F. Durand, and A. Torralba. A benchmark of computational models of saliency to predict human fixations. In *MIT Technical Report*, 2012.

[18] T. Judd, K. Ehinger, F. Durand, and A. Torralba. Learning to predict where humans look. In *IEEE ICCV*, 2010.

[19] W. Kienzle, F. A. Wichmann, M. O. Franz, and B. Schölkopf. A nonparametric approach to bottom-up visual saliency. In *NIPS*, 2007.

[20] S. S. Kruthiventi, K. Ayush, and R. V. Babu. Deepfix: A fully convolutional neural network for predicting human eye fixations. *IEEE TIP*, 26(9):4446–4456, 2017.

[21] A. Kucuk, J. M. Banda, and R. A. Angryk. A large-scale solar dynamics observatory image dataset for computer vision applications. *Scientific data*, 4:170096, 2017.

[22] J. Li, L. Y. Duan, X. Chen, T. Huang, and Y. Tian. Finding the secret of image saliency in the frequency domain. *IEEE TPAMI*, 37(12):2428–2440, 2015.

[23] J. Li, M. D. Levine, X. An, X. Xu, and H. He. Visual saliency based on scale-space analysis in the frequency domain. *IEEE TPAMI*, 35(4):996–1010, 2013.

[24] J. Li, Y. Tian, and T. Huang. Visual saliency with statistical priors. *IJCV*, 107(3):239–253, 2014.

[25] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille. The secrets of salient object segmentation. In *IEEE CVPR*, 2014.

[26] N. Liu, J. Han, T. Liu, and X. Li. Learning to predict eye fixations via multiresolution convolutional neural networks. *IEEE TNNLS*, 29(2):392–404, 2018.

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

[28] J. Pan, E. Sayrol, X. Giro-i Nieto, K. McGuinness, and N. E. O'Connor. Shallow and deep convolutional networks for saliency prediction. In *IEEE CVPR*, 2016.

[29] H. R.-Tavakoli, A. Borji, J. Laaksonen, and E. Rahtu. Exploiting inter-image similarity and ensemble of extreme learners for fixation prediction using deep features. *Neurocomputing*, 244, 2016.

[30] H. R. Tavakoli, E. Rahtu, and J. Heikkilä. Fast and efficient saliency detection using sparse sampling and kernel density estimation. In *Scandinavian Conference on Image Analysis*, pages 666–675. Springer, 2011.

[31] E. Vig, M. Dorr, and D. Cox. Large-scale optimization of hierarchical features for saliency prediction in natural images. In *IEEE CVPR*, 2014.

[32] W. Wang and J. Shen. Deep visual attention prediction. *IEEE TIP*, 27(5):2368–2378, 2018.

[33] C. Yang, L. Zhang, H. Lu, R. Xiang, and M. H. Yang. Saliency detection via graph-based manifold ranking. In *IEEE CVPR*, 2013.

[34] J. Zhang and S. Sclaroff. Saliency detection: A boolean map approach. In *IEEE ICCV*, pages 153–160, 2013.

[35] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of vision*, 8(7):32–32, 2008.