Spectral Transfer Guided Active Domain Adaptation For Thermal Imagery

Berkcan Ustun[1]   Ahmet Kagan Kaya[1,2]   Ezgi Cakir Ayerden[1,2]   Fazil Altinel[1]

[1]Research Center, Aselsan Inc., Turkiye

[2]Department of Electrical and Electronics Engineering, Middle East Technical University, Turkiye

{berkcanustun, kagankaya, eayerden, faltinel}@aselsan.com.tr

Abstract

The exploitation of visible spectrum datasets has led deep networks to show remarkable success. However, real-world tasks include low-lighting conditions which arise performance bottlenecks for models trained on large-scale RGB image datasets. Thermal IR cameras are more robust against such conditions. Therefore, the usage of thermal imagery in real-world applications can be useful. Unsupervised domain adaptation (UDA) allows transferring information from a source domain to a fully unlabeled target domain. Despite substantial improvements in UDA, the performance gap between UDA and its supervised learning counterpart remains significant. By picking a small number of target samples to annotate and using them in training, active domain adaptation tries to mitigate this gap with minimum annotation expense. We propose an active domain adaptation method in order to examine the efficiency of combining the visible spectrum and thermal imagery modalities. When the domain gap is considerably large as in the visible-to-thermal task, we may conclude that the methods without explicit domain alignment cannot achieve their full potential. To this end, we propose a spectral transfer guided active domain adaptation method to select the most informative unlabeled target samples while aligning source and target domains. We used the large-scale visible spectrum dataset MS-COCO as the source domain and the thermal dataset FLIR ADAS as the target domain to present the results of our method. Extensive experimental evaluation demonstrates that our proposed method outperforms the state-of-the-art active domain adaptation methods. The code and models are publicly available.[1]

1. Introduction

The latest state-of-the-art deep learning methods have led to a substantial improvement on computer vision and pattern recognition tasks such as classification and object detection with the use of RGB images [16, 20, 33, 34]. Deep models trained on massive RGB datasets e.g., ImageNet [9], MS-COCO [26], Pascal-VOC [11], etc. have shown considerable performance. Despite the progress of recent deep learning models trained on the visible spectrum images, low-lighting conditions prevent the most current models from performing competently on several real-world tasks. Since thermal cameras are more robust against these conditions, using them is advantageous for real-world applications such as military operations, security monitoring, autonomous driving, etc. However, large scale thermal datasets are not easily accessible to the public. Since the most advanced models require a massive amount of labeled data, models trained on thermal images struggle to perform as well as models trained on RGB images. One simple strategy to enhance the effectiveness of systems that use thermal imaging for classification and detection tasks is to take advantage of the additional information provided by visible spectrum images. Unfortunately, the models trained on visible spectrum datasets may perform poorly when the models tested on thermal images [3, 10, 18].

In order to overcome the performance drop on target thermal domain while utilizing the source visible spectrum domain, the domain shift between the visible and thermal spectrum images must be taken care of. Domain adaptation (DA) methods intend to learn a mapping from both domains to the same feature space. Unsupervised domain adaptation (UDA) methods are based on the assumption that labels for target domain are not available [14, 35, 41]. Labeling all of the data is costly in terms of time and human resources. However, limitation in semantic annotations is more flexible in real world applications since a reasonable portion of the target samples can be annotated in consideration of the labeling costs and budget. In this sense, the concept of active learning is useful for determining which portion of the data should be annotated. Most active learning methods have recently concentrated on developing a query function that measures how representative and informative a sample is based primarily on uncertainty [39]. However, domain shift between source and target samples hinders classical

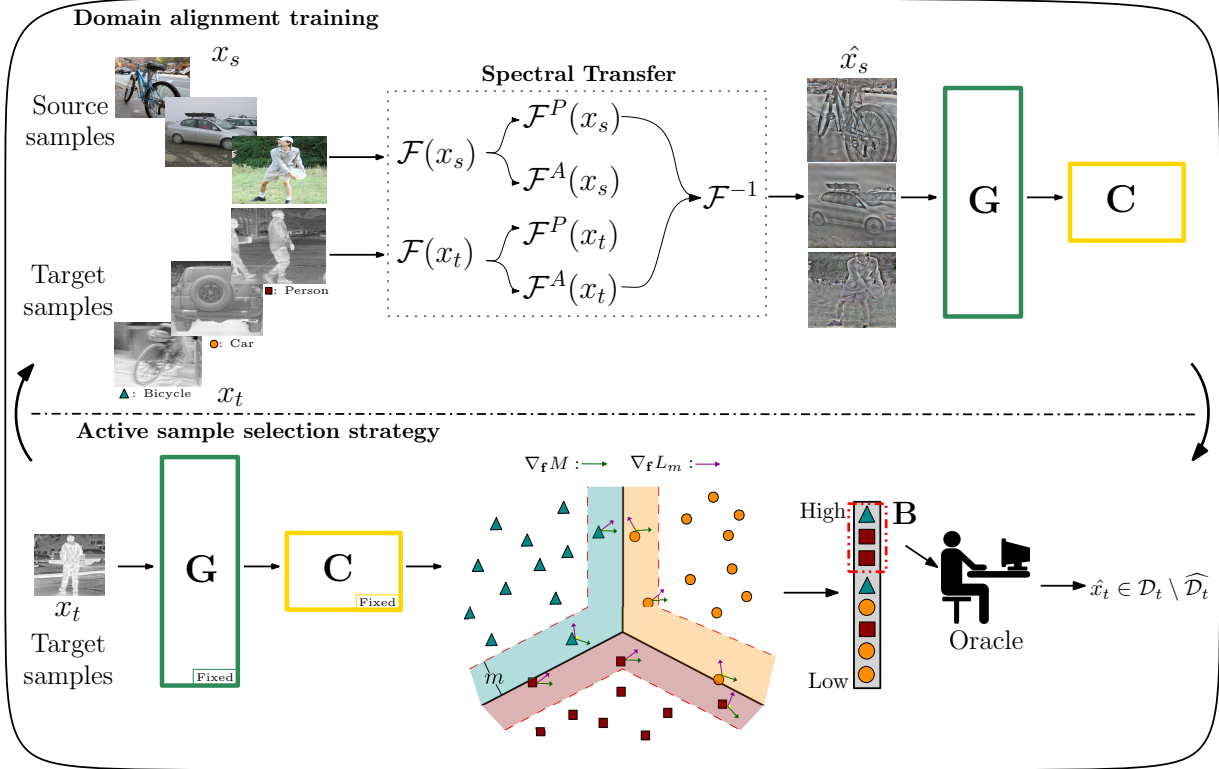---

[1]https://github.com/avaapm/STGADA

Figure 1. An overview of our proposed spectral transfer guided active domain adaptation (STGADA) method. Source samples are subjected to spectral transfer with target samples. The amplitude component of the Fourier transform of the source samples $\mathcal{F}^A(x_s)$ are exchanged with that of target samples $\mathcal{F}^A(x_t)$. We train our feature extractor $\mathbf{G}$ and classifier $\mathbf{C}$ with the transformed source samples $x_{s \to t}$. Then, we proceed to active sample selection. We use the fixed feature extractor $\mathbf{G}$ and the classifier $\mathbf{C}$ to sample from the target dataset. In the feature space representation of the target samples, the query function measures how close a target sample to the decision boundary M. In addition, our sampling strategy considers how similar the direction of the gradient of the loss term $\nabla_{\mathbf{f}} L_m$ and that of margin term $\nabla_{\mathbf{f}} M$. Oracle annotates target samples of labeling budget $\mathbf{B}$ (dashed red box) to use in domain alignment training. *Best viewed in color.*

active learning query methods from selecting the samples which could increase the performance. The query methods that exploit the prediction of the task models trained only on source domain tend to sample some irrelevant target samples since most of the target data considered uncertain by the model. Therefore, active learning in consideration of domain shift problem is an important research area [13, 40, 46].

Recent state-of-the-art active domain adaptation methods consider the domain gap between the source and target datasets from two perspectives. Approaches such as [32], [40], and [45] focus on conducting an explicit domain adaptation to align the feature spaces of both domains. On the other hand, [13] and [46] seek a solution to the problem of active domain adaptation without concerning explicit alignment of feature spaces. When the domain gap is considerably large as in the visible-to-thermal task, we may conclude that the methods without domain adaptation cannot achieve their full potential. In other words, selected target samples by the query functions become efficient in the case of a moderate domain shift. In that case, implementing

additional domain adaptation algorithms on top of models does not yield a significant improvement. On the contrary, additional reduction of the domain gap may improve the performance in the RGB-to-thermal task since the domain gap is considerably large.

In this paper, we propose an active domain adaptation method to select the most informative samples while aligning visible and thermal domains as described in Sec. 3. We employ Select-by-Distinctive-Margin (SDM) [46] as our base method. Based on our experimental evaluations in Sec. 4, SDM is more calibrated and achieves more successful results compared to the state-of-the-art active domain methods. Although SDM and other domain alignment-free active domain adaptation methods have achieved successful results, additional reduction of the domain gap may improve the performance in the RGB-to-thermal task since the domain gap is large compared to the domain gap in classical domain adaptation problems. In that regard, we apply a domain adaptation method to the SDM [46] to achieve the promised performance improvement.

Domain adaptation algorithms come as a module imple-

mented additionally to a backbone which is used to extract features. Those modules are generally a few layered deep networks with complicated loss functions or regularization terms [14, 27, 35]. Apart from popular domain adaptation approaches, recently, using frequency domain information in order to reduce the domain gap has become popular in terms of the simplicity of manipulating the frequency characteristics of the samples without increasing the complexity of the baseline model [25, 47]. [47] proposes that interchanging the low frequency characteristics of the source and target samples reduces the domain gap. This method takes advantage of simplicity by applying the algorithm to the training samples just before they are input to the base model. Interchanging the low frequency components of source and target samples can be implemented as a transformation to the source samples that are used in the training of SDM [46]. With these in mind, we propose a spectral transfer guided active domain adaptation (STGADA) method in order to achieve higher performance in thermal domain via reducing the domain gap between visible and thermal IR spectra (Fig. 1). Spectral transfer is applied to the source domain samples by implementing the Fourier domain adaptation (FDA) [47] during training. The low-frequency components of the labeled source samples are changed with those of unlabeled target samples. This transfer reduces the domain gap between source and target domains. The transformed source samples are then fed to the backbone network. We sample from target thermal dataset according to a query function which uses the output of the fixed classifier i.e., the prediction scores of the overall model. This sampling process takes place in the specified epochs during training. Therefore, the training of the network continues after the sampling steps.

During training and testing our approach, we exploit the large scale visible spectrum dataset MS-COCO [26] and thermal spectrum dataset FLIR ADAS [17]. Moreover, our method does not require paired samples of RGB and thermal images. We conduct extensive qualitative and quantitative analysis in order to evaluate our method. Sec. 4 demonstrates that our approach outperforms the state-of-the-art methods with improving our base model. Fig. 3 depicts that given a model without domain alignment or active sampling i.e, the source only model, the calibration line is far from ideal. The union of predicted samples within a confidence interval does not yield proportionate accuracy. In contrast, our proposed model STGADA obtains more calibrated results leading to a narrower gap in the reliability diagram. Moreover, class imbalance is important to address since real world applications include imbalanced classes [23]. Experimental studies in Sec. 4 show that our model obtains more balanced performance than the state-of-the-art methods since those active domain adaptation methods tend to over-classify the majority category.

Our contributions are summarized as follows:

- We propose a simple yet efficient spectral transfer guided active domain adaptation approach for thermal IR spectrum. In order to efficiently reduce the domain gap between source and target domains, we employ a spectral transfer algorithm.

- We conduct extensive analysis in order to demonstrate the efficiency of our method in the domain adaptation setting where MS-COCO [26] is RGB source dataset and FLIR ADAS [17] is thermal target dataset. The results show that our method outperforms state-of-the-art active domain adaptation models by selecting more informative target thermal samples.

2. Related Work

Deep learning methods have achieved promising results on computer vision tasks by using RGB images. However, real-world conditions such as low-lighting challenge these models. Therefore, recent studies [3,19,21,22,36,42] investigated performance on the classification and object detection tasks in relation to the consequences of combining visible spectrum and thermal image modalities. Motivated by these studies, in this work, we take advantage of supplementary information provided by visible spectrum images to enhance classification performance on thermal imagery without requiring RGB-to-thermal image pairings since large-scale thermal datasets are not publicly available. To evaluate the effectiveness of combining RGB and thermal images, we propose an active domain adaption approach.

Active learning algorithms choose the samples to annotate with the semantic annotations rather than having access to the labels beforehand [38]. In order to choose the most disagreeing samples, query by committee approaches exploited many classifiers and measure their discrepancy on the task on unlabeled data using methods like low-dimension projections [15]. Query by uncertainty methods employed classification margin as a selection criteria [4], multi-class classification margin [24], learning loss [48], number of false-positive or false-negative pixels in images [2], discrepancy between two auxiliary classifiers [7]. To choose a collection of unique samples, representativeness approaches often utilized clustering or core-set selection [1, 37, 39]. Although, these methods achieved successful results on active learning problem , none of them considers the potential domain shift between labeled and unlabeled data.

Domain adaptation, in other words, the transfer of models trained on labeled source domains to unlabeled or partially labeled target domains, has drawn a lot of interest [5, 8, 14, 28]. Ganin et al. [14] proposed a gradient reversal layer to align source and target domains using the feature encoder which is trained to maximize the loss of the domain

discriminator while minimizing the loss of the classifier. Saito et al. [35] proposed to reduce the gap by employing adversarial learning with two task-specific classifiers and a feature extractor to align source and target domains. Besides from popular adversarial domain adaptation methods, using frequency information has gained interest to achieve distribution alignment. Yang et al. [47] proposed Frequency Domain Adaptation (FDA) method to overcome domain shift problem by simply replacing the low frequency component of a source sample with a target sample since the low frequency component of the samples provides the domain invariant properties. Therefore [47] introduces Frequency Domain Adaptation (FDA). Recently, Akkaya et al. [3] proposed a pseudo-labeling guided unsupervised domain adaptation method by using large-scale RGB dataset MS-COCO as source domain and thermal dataset FLIR ADAS as target domain. Although adversarial and frequency-level domain adaptation methods have achieved successful experimental results, active domain adaptation can boost the performance of domain adaptation by labeling small portion of unlabeled target dataset. Since a reasonable portion of the target samples can be annotated in consideration of the labeling costs and budget, domain adaptation can be combined with active learning to improve performance on RGB-to-thermal domain adaptation problem.

Active domain adaptation has gained increasing attention in recent years due to its potential to improve model performance in real-world scenarios. Su et al. [40] proposed active adversarial domain adaptation (AADA) method to employ adversarial training of a feature generator and a domain discriminator for domain alignment and present a sample selection approach based on variety and uncertainty of target samples. Using a new sample selection methodology known as CLUE, Prabhu et al. [31] suggested clustering target data embeddings weighted by uncertainty and choosing nearest neighbors to the cluster centroids for annotation while aligning domains using [14]. Fu et al. [13] proposed transferable query selection (TQS) method in their sample selection strategy which consists of three factors: discrepancy among a group of five classifiers, margin of projected class probabilities for a target sample, and sample diversity from the source domain via a domain discriminator. To choose the most informative subset from the target domain to annotate, Rangwani et al. [32] proposed S^{3}VAADA, which provides a submodular set-based information criterion that consists of three scores: virtual adversarial pairwise score, diversity score, and representativeness score. Based on energy-based models, Xie et al. [45] find target samples which are particularly distinctive to the target distribution by taking use of free energy biases. More recently, Xie et al. proposed Select-by-Distinctive-Margin (SDM) method which comprises of a maximum margin loss and a margin sampling algorithm. Using greatest margin

loss, SDM chooses target samples based on their relationship to a few hard instances from the source domain which are near to the decision boundary between source and target domain. Although SDM presents remarkable results, there is still room for improving performance on the tasks which have large domain gap by applying domain alignment.

Our approach differs from the past approaches that it manages the domain shift reduction, which is crucial in the case of the transfer of information from visible to thermal spectrum, without being overly complicated. To the best of our knowledge, there is no active domain adaptation study in the literature of thermal image classification. While the SDM [46] employs no extra classifiers for query by committee or complex architectures for aligning the distribution for the source and target domains, this lack of alignment leaves a space for improvement by reducing the domain gap. In addition, our choice of FDA [47] introduces no extra learnable parameters although it shows proper capability on dealing with the domain shift problem.

3. Proposed Method

Our proposed spectral transfer guided active domain adaptation method is illustrated in Fig. 1.

First, we change low-frequency components of the labeled source samples with those of unlabeled target samples. We take Fourier transform $\mathcal{F}$ of source and target samples. The output of the Fourier transform divides into two; phase and amplitude components. Then, the amplitude component of the target sample is multiplied with a mask M_β whereas that of the source sample is multiplied with the complement of that mask $(1 - M_\beta)$. These amplitude components are summed to generate the new amplitude component. With this process, we control the portion size of the low frequency part that is exchanged. The phase component of the source sample and the newly generated mixed amplitude component are subjected to inverse Fourier transform $\mathcal{F}^{-1}$. In this way, a spectral transfer which aids to reduce the domain gap between two datasets occurs between source and target samples. The transformed source samples are then fed to the feature extractor $\mathbf{G}$. A linear classifier $\mathbf{C}$ is trained along with the $\mathbf{G}$.

We sample from the target dataset according to a query function $Q(\cdot)$ which uses the output of the fixed $\mathbf{C}$ in other words, the prediction scores of the overall model. The query function $Q(\cdot)$ also uses the backward gradient of the loss function of the backbone network. Since target samples have no labels, this gradient is estimated for target samples. Once the selected target samples are labeled by oracle, we utilize the selected target samples in training of $\mathbf{G}$ and $\mathbf{C}$ with their labels. This sampling process takes place in the specified epochs during training. Therefore, the training of the network continues after the sampling steps.

In the inference, the target samples pass through the

backbone network which consists of the feature extractor **G** and the classifier **C**. Thus, inference with our proposed approach is remarkably simple.

A general definition of active domain adaptation, spectral transfer, and spectral transfer guided active domain adaptation procedures of our proposed approach are described in Sec. 3.1, Sec. 3.2, and Sec. 3.3, respectively.

3.1. Active Domain Adaptation

In active domain adaptation, the source domain has n_s labeled samples which is represented as $\mathcal{D}_s = \{(x_s^i, y_s^i)\}_{i=1}^{n_s}$ with samples x_s^i and their semantic annotations $y_s^i \in \{1, 2, \ldots K\}$ where K is the number of classes. The target domain has n_t unlabeled samples which is denoted as $\mathcal{D}_t = \{(x_t^j)\}_{j=1}^{n_t}$. In addition, we describe a labeled target set which is initially an empty set $\varnothing$, as $\widehat{\mathcal{D}_t}$. The unlabeled target data is sampled several times. The annotators assign their labels to the selected target data, $\hat{x}_t \in \mathcal{D}_t \setminus \widehat{\mathcal{D}_t}$. The model can be trained with the new labeled set, $\widehat{\mathcal{D}_t} \cup \mathcal{D}_s$, and then it will be exploited to sample new unlabeled data from the set $\mathcal{D}_t \setminus \widehat{\mathcal{D}_t}$. Given a budget **B**, this procedure continues until the annotated number of target samples achieves the predefined budget $|\widehat{\mathcal{D}_t}| = \mathbf{B}$.

3.2. Spectral Transfer

In this subsection, we elaborate the spectral transfer procedure exploited in our proposed method (dashed box in the upper part of Fig. 1).

For the Fourier transform $\mathcal{F}$ of an RGB image, $\mathcal{F}^A, \mathcal{F}^P : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^{H \times W \times 3}$ represent the amplitude and phase components, respectively. For simplicity of explanation, Fourier transform $\mathcal{F}$ of a single channel image x is:

$$\mathcal{F}(x)(u, v) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} x(h, w) e^{-j2\pi(\frac{h}{H}u + \frac{w}{W}v)}. \quad (1)$$

This can be efficiently implemented by FFT algorithm proposed in [12]. In order to map frequency domain signals back to image space, we perform inverse Fourier transform $\mathcal{F}^{-1}$. A mask M_β is defined whose value is zero except for the region in the center $(h, w) \in [-\beta H : \beta H, -\beta W : \beta W]$ where $\beta \in (0, 1)$ with the assumption that the center is $(0, 0)$. We can define the spectral transfer from $x_t \in \mathcal{D}_t$ to $x_s \in \mathcal{D}_s$ as:

$$\hat{x}_s = \mathcal{F}^{-1}([M_\beta \circ \mathcal{F}^A(x_t) + (1 - M_\beta) \circ \mathcal{F}^A(x_s), \mathcal{F}^P(x_s)]) \quad (2)$$

where the low-frequency component of the amplitude of the source sample is replaced by that of the target sample. The operation $\circ$ denotes the point-wise multiplication of the image and the mask matrices. The phase component of the source image is unchanged. This altered spectral signal is then mapped back to the image domain. The image $\hat{x}_s$ has the same content as x_s while it constitutes the style and appearance of the sample x_t from $\mathcal{D}_t$.

3.3. Spectral Transfer Guided Active Domain Adaptation

In the task of active domain adaptation, all labeled data $\mathcal{D}_s \cup \widehat{\mathcal{D}_t}$ is used. This situation arises the problem of bias towards areas in the source domain where data samples are densely placed. Class imbalance problem in real-world applications amplifies this bias. Therefore, a network which is trained using all of the training data may be overfit to the abundant class. Moreover, this bias prevents the $Q(\cdot)$ from selecting informative target samples. Arising from this fact, a categorical-wise margin loss with its selective property can be employed to supervise the network. The proportional contribution of the source samples to the backward gradient concerning their margin size helps the network to focus on samples that have different difficulties. The consideration of uncertain source samples helps network to prevent the overfitting to the abundant class. Furthermore, we employ a max-logit regularizer to ensure that the network always assigns large scores to the prediction on ground truth class. Assigning large scores to the prediction on the ground truth class is important for calibration. We expect that the accuracy of the union of samples within a confidence level is similar with that confidence for a calibrated model. Concerning calibration, confidence score needs to be as large as accuracy. Hence, considering the spectral transfer in Sec. 3.2 and adaptive margin loss, the overall loss function can be denoted as:

$$L_m(\hat{x}_s, y_s) = \sum_{i \neq j} [\gamma_i[m - \mathbf{C}(\mathbf{G}(\hat{x}_s))_j + \mathbf{C}(\mathbf{G}(\hat{x}_s))_i]_+$$
$$- \mathbf{C}(\mathbf{G}(\hat{x}_s))_j] \quad (3)$$
$$\gamma_i = 1 - \frac{\mathbf{C}(\mathbf{G}(\hat{x}_s))_j - \mathbf{C}(\mathbf{G}(\hat{x}_s))_i}{m}$$

where $[x]_+$ is the operation $max(0, x)$, the subscripts i and j denotes the i-th and j-th entry of vectors, respectively. m is a hyper-parameter to control the margin width. Here, only the samples which have a confidence score of ground-truth class close to the classification scores of other classes can contribute to the backward gradient. In addition, we control the contribution by using γ_i which is proportional to the similarity between scores of the ground truth class and those of other classes.

While this loss function enhances the spaces between different categorical clusters, we can focus on the target samples that are placed near smaller gaps between categorical clusters in the feature space. Furthermore, ensuring the gradient from loss and the margin sampling term present similar orientations in the feature space will guarantee fast convergence to a robust state. Let the margin term be defined as:

$$M(x_t) = (1 - (\mathbf{p}_{1^*} - \mathbf{p}_{2^*})) \quad (4)$$

where;

$$\mathbf{p} = \mathbf{softmax}(\mathbf{C}(\mathbf{G}(x_t))) \quad (5)$$

and subscript 1^* and 2^* denotes the maximum and second maximum prediction scores.

By considering these, the overall query function for a sample x can be denoted as:

$$Q(x) = M(x) + \lambda \langle \nabla_{\mathbf{f}} L_m(x, y), \nabla_{\mathbf{f}} M(x) \rangle \quad (6)$$

where $\mathbf{f} = \mathbf{G}(x)$. In Eq. (6), λ is a hyper-parameter and operation $\langle \cdot, \cdot \rangle$ defines the Cosine Similarity. Since we do not have access to the label of the target samples, the gradient from the loss term cannot be calculated, but can be estimated:

$$\nabla_{\mathbf{f}} \hat{L}_m(x, y) = \mathbf{p}_{1^*} \nabla_{\mathbf{f}} L_m(x, 1^*) + \mathbf{p}_{2^*} \nabla_{\mathbf{f}} L_m(x, 2^*). \quad (7)$$

With the query function $Q(x)$, we can select the most informative target samples that are placed near confused decision boundaries and ensure fast convergence. Similar with the categorical margin loss in Eq. (3), the query function $Q(x)$ considers the uncertain samples.

4. Experiments

We carry out comprehensive analyses and compare our proposed method with a number of state-of-the-art unsupervised domain adaption techniques.

4.1. Datasets

We employed the RGB-to-thermal domain adaptation setting which is recently proposed by Akkaya et al. [3]. The setting consists of two datasets, namely, FLIR ADAS [17] as a thermal dataset and MS-COCO [26] as an RGB dataset.

FLIR ADAS [17] contains both RGB images and thermal images. We utilized only the thermal images in our experiments. As in [3], we used three classes from FLIR ADAS dataset: bicycle, car, and person. MS-COCO [26] is a publicly available large-scale visible spectrum dataset that contains the same classes as FLIR ADAS [17]. Therefore, MS-COCO [26] is considered visible spectrum dataset in our experiments in parallel with [3].

The images in both datasets originate from the square bounding box annotations of the objects i.e, bicycle, car, person. The extracted object images are resized to 224×224. Consequently, the thermal dataset includes 4,137 samples of bicycles, 43,734 samples of cars, and 26,294 samples of person images. The RGB dataset consists of 5,732 samples of bicycles, 38,453 samples of cars, and 209,162 samples of person images.

4.2. Implementation Details

We followed the training procedure of [46] in our experiments. We used ResNet-50 pre-trained on ImageNet feature extractor $\mathbf{G}$ for all approaches. The FDA [47] method is implemented on top of our backbone as a spectral transformation. During the training of our method, we set the batch size to 32. The number of epochs was determined to be 50.

Table 1. Per-class classification performance comparison.

Method	Bicycle	Car	Person	Average
Source only	69.89	83.89	86.52	80.10
Random	63.91	96.73	97.87	86.17
Entropy [44]	72.41	97.38	96.49	88.76
S^{3}VAADA [32]	56.78	72.13	**98.77**	75.89
EADA [45]	74.18	94.41	96.52	88.37
AADA [40]	81.15	97.81	94.65	91.20
CLUE [31]	86.21	96.95	91.75	92.30
SDM [46]	88.05	98.35	96.51	94.30
TQS [13]	90.34	97.93	96.41	94.89
STGADA (Ours)	**90.57**	**98.96**	96.53	**95.35**

AdaDelta optimization algorithm is employed to update the parameters. The learning rate was set to 0.5. The parameters λ and β were chosen as 0.001 and 0.033, respectively. The margin size m was selected as 1. Throughout the experiments, we perform 5 rounds of active sampling. 2% of target samples are labeled at every round. Following the standard active domain adaptation approaches [13, 40, 46], we simulate oracle annotations by using the ground truth. Labeled target data are extracted from unlabeled target set and added to source set. For the first 10 epochs, the training proceeded with the initial source data. In epochs 10, 12, 14, 16, and 18, we perform our sample selection strategy.

We implemented our proposed method using PyTorch framework [30]. Implementation details, models, and the code were made publicly available at https://github.com/avaapm/STGADA.

4.3. Results

In the experiments, the visible spectrum is chosen as the source domain, and the thermal IR spectrum is chosen as the target domain. As a general practice in domain adaptation, we denote *source only* as the target dataset performance of a model trained with only the source dataset. Performance of *source only* model serve as baseline for the lower bound performance.

Quantitative Analysis. We compare our proposed method STGADA with several state-of-the-art active domain adaptation methods, namely Active Adversarial Domain Adaptation (AADA) [40], Clustering Uncertainty-weighted Embeddings (CLUE) [31], Energy-based Active Domain Adaptation (EADA) [45], Submodular Subset Selection for Virtual Adversarial Active Domain Adaptation (S^{3}VAADA) [32], Transferable Query Selection (TQS) [13], and Select by Distinctive Margin (SDM) [46]. More-

● : Bicycle   ● : Car   ● : Person

□ : Selected samples

(a) Source Only   (b) SDM
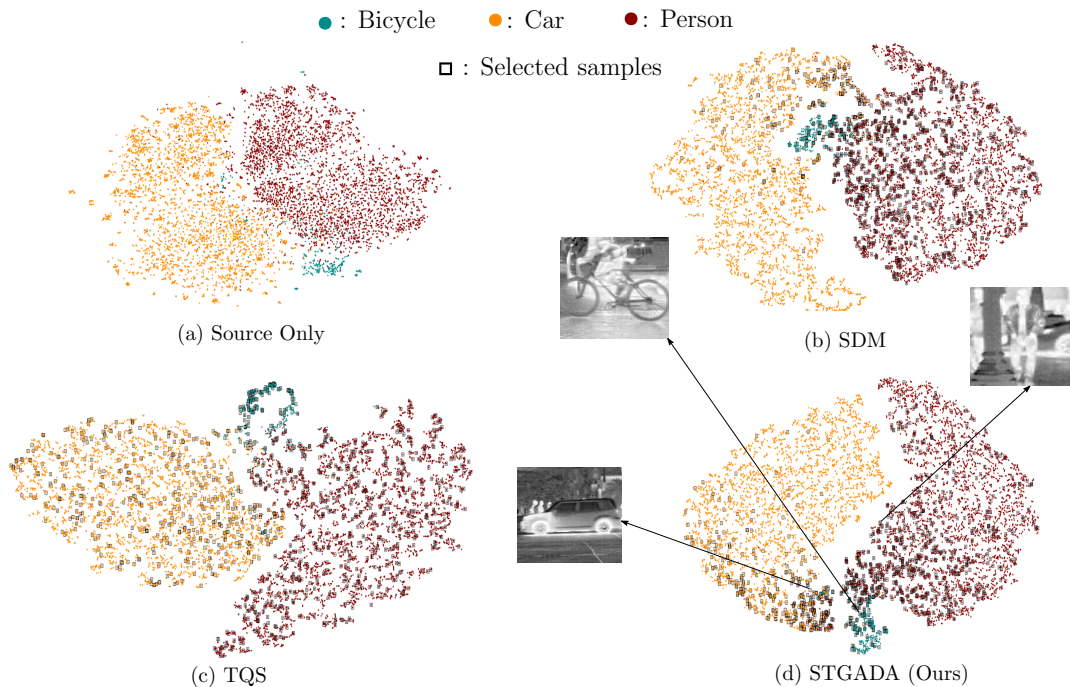
(c) TQS   (d) STGADA (Ours)

Figure 2. The t-SNE [43] visualization of network activations on target thermal domain generated by source only model (a), our base method SDM [46] (b), TQS [13] (c), and our proposed method STGADA (d). We visualize one sample per each class selected by our proposed method STGADA. *Best viewed in color.*

over, we compare our proposed method with baseline active learning sampling strategies: random sampling and entropy sampling [44]. There are no published findings for our dataset in these studies since these approaches do not take domain adaptation into account for thermal datasets. As a result, we trained and evaluated each of these approaches using our dataset. We also conduct an analysis on the calibration of the source only model, SDM [46], TQS [13] and our proposed model STGADA. The Expected Calibration Error (ECE) [29] values are reported in Tab. 2 and Fig. 3 depicts the reliability diagram of the models.

Table 2. Expected Calibration Error (ECE) [29] results. The lower the better.

Method	ECE
Source only	0.0856
TQS [13]	0.0122
SDM [46]	0.0041
STGADA (Ours)	0.0044

Per-class classification accuracy results are reported in Tab. 1. The results show that our proposed method outperforms the state-of-the-art models. According to per-class accuracy, our method achieves the best performances on the classes *bicycle* and *car*. Although S^{3}VAADA [32] performs

well for the *person* class, the accuracy of that model on the other classes shows that the model overfits to *person* class, which is an abundant class in the source domain dataset. On the other hand, our approach achieves balanced performance for all classes despite the fact that the source dataset is imbalanced. Our method also increases the performance of the base method SDM [46] in the *bicycle* class, which has very few samples in the source dataset. It is important to underline this situation because real-world applications usually include imbalanced classes [6, 23].

Tab. 2 shows the Expected Calibration Error (ECE) [29] results. If a classifier's observed accuracy matches with its mean confidence, the classifier is considered to be calibrated [29]. From the results, we can conclude that the source only model has the highest ECE, which states that the source only model is not calibrated well. In terms of calibration, our model STGADA and SDM [46] outperforms TQS [13] by obtaining much less ECE. In parallel to ECE results, Fig. 3 demonstrates how far the source only model is from the perfect calibration line. In other words, the source only model cannot achieve the expected accuracy concerning obtained confidence scores. TQS [13] also exhibits unwanted peaks and gaps stating that the model is not calibrated enough. On the other hand, our proposed model STGADA and SDM [46] manage to achieve proportional accuracy to confidence scores that the gaps in the reliability diagrams are smaller than those of the source only model

(a) Source only   (b) SDM
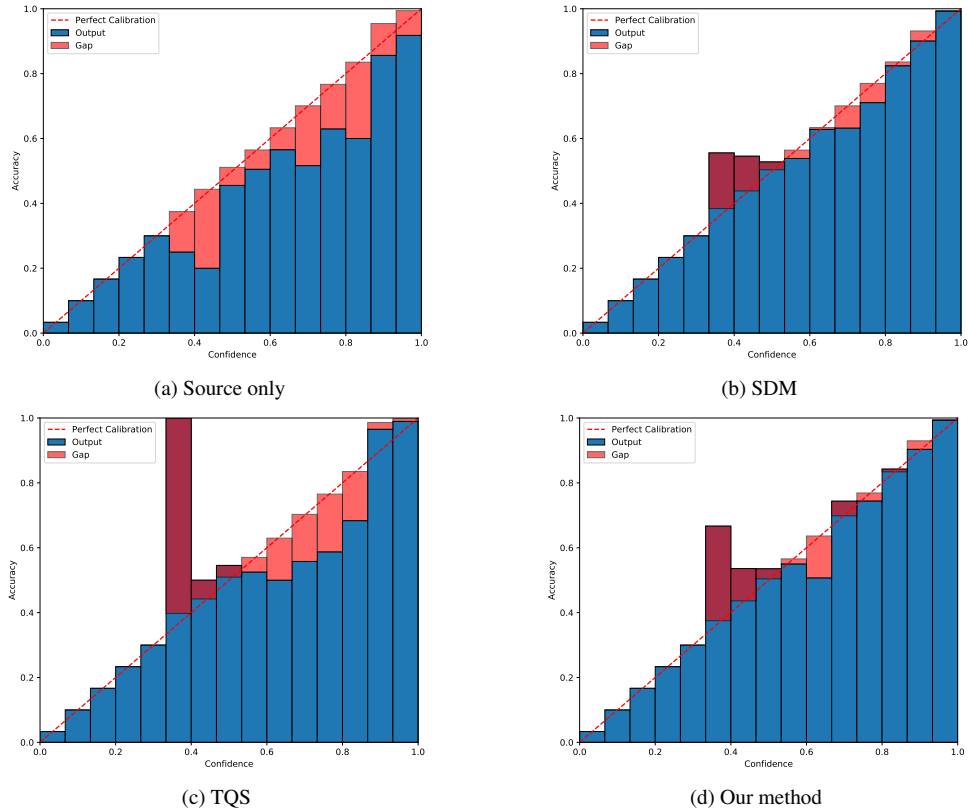
(c) TQS   (d) Our method

Figure 3. Visualization of reliability diagrams [29] of source only model (a), SDM [46] (b), TQS [13] (c), and our proposed method STGADA (d). Given target domain images, our proposed method STGADA obtains more calibrated results compared to the source only model and TQS [13] while showing comparable calibration gap with SDM [46].

and TQS [13]. Since the calibration of the SDM [46] is better than that of the TQS [13], we have chosen the SDM [46] as the base model for our approach. Furthermore, the fact that TQS [13] has a high level of complexity with multiple classifiers and a discriminator underlines our base method decision of SDM [46]. Consequently, our model is more calibrated than TQS [13] and comparable with SDM [46] in addition to achieving higher accuracy.

Qualitative Analysis We present the visualization of the feature representation on the target thermal IR domain with t-SNE [43] for qualitative analysis in Fig. 2. The features originating from the source only model cannot be discriminated, while our base method SDM [46] and TQS [13] discriminate the samples from different classes. On the other hand, our proposed model STGADA can isolate the samples from the *bicycle* better than those models. The improvement in the discrimination of the samples from *bicycle* reflects itself in the fact that our model achieved the best accuracy in *bicycle* class. Furthermore, we visualized one selected sample from each class, *bicycle*, *car*, and *person* in Fig. 2. Our proposed approach tends to select the most uncertain samples that rely upon near the decision boundaries. As we can see from the Fig. 2, samples, which are close to different categorical clusters are selected. The vi-

sualized examples also confirm this criteria. We can see that the samples from the classes *bicycle* and *car* contain a person object, and the sample from the class *person* contains a bicycle object. This situation makes those samples ambiguous for the trained network, hence, the query function selects those samples.

5. Conclusion

In this study, we propose a spectral transfer guided active domain adaptation method in order to examine the efficiency of combining visible spectrum and thermal imagery modalities by taking advantage of active domain adaptation. We perform spectral transfer to the source samples with the target samples for additional alignment of two domains. The low frequency component of source samples are exchanged with those of target samples. This extra alignment leads our model to achieve improved performance since the domain shift in RGB-to-thermal task is considerably large. To present our results, we employed the large scale RGB dataset MS-COCO as the source domain and thermal dataset FLIR ADAS as the target domain. Quantitative and qualitative analyses demonstrate that our proposed approach performs better than state-of-the-art active domain adaptation methods by reducing the domain gap between RGB and thermal domains.

References

[1] Sharat Agarwal, Himanshu Arora, Saket Anand, and Chetan Arora. Contextual diversity for active learning. In *ECCV*, 2020. 3

[2] Hamed H Aghdam, Abel Gonzalez-Garcia, Joost van de Weijer, and Antonio M López. Active learning for deep detection neural networks. In *ICCV*, 2019. 3

[3] Ibrahim Batuhan Akkaya, Fazil Altinel, and Ugur Halici. Self-training guided adversarial domain adaptation for thermal imagery. In *CVPRW*, 2021. 1, 3, 4, 6

[4] Maria-Florina Balcan, Andrei Broder, and Tong Zhang. Margin based active learning. In *COLT*, 2007. 3

[5] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. *NeurIPS*, 2016. 3

[6] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 2018. 7

[7] Jae-Won Cho, Dong-Jin Kim, Yunjae Jung, and In-So Kweon. Mcdal: Maximum classifier discrepancy for active learning. In *IEEE transactions on neural networks and learning systems*, 2021. 3

[8] Debasmit Das and CS George Lee. Unsupervised domain adaptation using regularized hyper-graph matching. In *ICIP*, 2018. 3

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 1

[10] Chaitanya Devaguptapu, Ninad Akolekar, Manuj M Sharma, and Vineeth N Balasubramanian. Borrow from anywhere: Pseudo multi-modal object detection in thermal imagery. In *CVPR*, 2019. 1

[11] Mark Everingham, Luc Gool, Christopher K. Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *IJCV*, 2010. 1

[12] M. Frigo and S.G. Johnson. Fftw: an adaptive software architecture for the fft. In *ICASSP*, 1998. 5

[13] Bo Fu, Zhangjie Cao, Jianmin Wang, and Mingsheng Long. Transferable query selection for active domain adaptation. In *CVPR*, 2021. 2, 4, 6, 7, 8

[14] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015. 1, 3, 4

[15] Ran Gilad-Bachrach, Amir Navot, and Naftali Tishby. Query by committee made real. In *NeurIPS*, 2006. 3

[16] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014. 1

[17] F. A. Group. Flir thermal dataset for algorithm training. https://www.flir.com/oem/adas/adas-dataset-form/. 3, 6

[18] Dayan Guan, Xing Luo, Yanpeng Cao, Jiangxin Yang, Yanlong Cao, George Vosselman, and Michael Ying Yang. Unsupervised domain adaptation for multispectral pedestrian detection. In *CVPRW*, 2019. 1

[19] Tiantong Guo, Cong Phuoc Huynh, and Mashhour Solh. Domain-adaptive pedestrian detection in thermal images. In *ICIP*, 2019. 3

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[21] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baseline. In *CVPR*, 2015. 3

[22] Shu Wang Jingjing Liu, Shaoting Zhang and Dimitris Metaxas. Multispectral deep neural networks for pedestrian detection. In *BMVC*, 2016. 3

[23] Justin M Johnson and Taghi M Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 2019. 3, 7

[24] Ajay J. Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *CVPR*, 2009. 3

[25] Zhaowen Li, Xu Zhao, Chaoyang Zhao, Ming Tang, and Jinqiao Wang. Transfering low-frequency features for domain adaptation. In *ICME*, 2022. 3

[26] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and Larry Zitnick. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1, 3, 6

[27] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, 2018. 3

[28] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu. Transfer feature learning with joint distribution adaptation. In *ICCV*, 2013. 3

[29] Mahdi Pakdaman Naeini, Gregory Cooper, and Milos Hauskrecht. Obtaining well calibrated probabilities using bayesian binning. *AAAI*, 2015. 7, 8

[30] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *NeurIPS*, 2019. 6

[31] Viraj Prabhu, Arjun Chandrasekaran, Kate Saenko, and Judy Hoffman. Active domain adaptation via clustering uncertainty-weighted embeddings. In *CVPR*, 2021. 4, 6

[32] Harsh Rangwani, Arihant Jain, Sumukh K Aithal, and R Venkatesh Babu. S3vaada: Submodular subset selection for virtual adversarial active domain adaptation. In *ICCV*, 2021. 2, 4, 6, 7

[33] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016. 1

[34] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015. 1

[35] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018. 1, 3, 4

[36] Philip Saponaro, Scott Sorensen, Abhishek Kolagunda, and Chandra Kambhamettu. Material classification with thermal imagery. In *CVPR*, 2015. 3

[37] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. In *ICLR*, 2018. 3

[38] Burr Settles. From theories to queries: Active learning in practice. In *AISTATS*, 2011. 3

[39] Samarth Sinha, Sayna Ebrahimi, and Trevor Darrell. Variational adversarial active learning. In *CVPR*, 2019. 1, 3

[40] Jong-Chyi Su, Yi-Hsuan Tsai, Kihyuk Sohn, Buyu Liu, Subhransu Maji, and Manmohan Chandraker. Active adversarial domain adaptation. In *WACV*, 2020. 2, 4, 6

[41] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *CVPR*, 2017. 1

[42] Engin Uzun, Ahmet Anıl Dursun, and Erdem Akagündüz. Augmentation of atmospheric turbulence effects on thermal adapted object detection models. In *CVPRW*, 2022. 3

[43] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 2008. 7, 8

[44] Dan Wang and Yi Shang. A new active labeling method for deep learning. In *IJCNN*, 2014. 6, 7

[45] Binhui Xie, Longhui Yuan, Shuang Li, Chi Harold Liu, Xinjing Cheng, and Guoren Wang. Active learning for domain adaptation: An energy-based approach. In *AAAI*, 2022. 2, 4, 6

[46] Ming Xie, Yuxi Li, Yabiao Wang, Zekun Luo, Zhenye Gan, Zhongyi Sun, Mingmin Chi, Chengjie Wang, and Pei Wang. Learning distinctive margin toward active domain adaptation. In *CVPR*, 2022. 2, 3, 4, 6, 7, 8

[47] Y. Yang and S. Soatto. Fda: Fourier domain adaptation for semantic segmentation. In *CVPR*, 2020. 3, 4, 6

[48] Donggeun Yoo and In So Kweon. Learning loss for active learning. In *CVPR*, 2019. 3