ENHANCING VEHICLE DETECTION ACCURACY UNDER LOW LIGHT CONDITIONS USING INFRARED IMAGES

Shivom Bhargava
M. Tech student,
Civil Engineering Department
Indian Institute of Technology Kanpur, India

Pranamesh Chakraborty, Corresponding Author
Assistant Professor,
Civil Engineering Department
Indian Institute of Technology Kanpur, India

Varsha Ravi-Chandra Mouli
Ph.D student,
Computer Science Department
Iowa State University, USA

Word Count: 6243 words + 1 table(s) × 250 = 6493 words

Submission Date: March 28, 2023

ABSTRACT

Vehicle detection accuracy is fairly accurate in good-illumination conditions but susceptible to poor detection accuracy under low-light conditions. The combined effect of low-light and glare from vehicle headlight or tail-light results in misses in vehicle detection more likely by state-of-the-art object detection models. However, thermal infrared images are robust to illumination changes and are based on thermal radiations. Recently, Generative Adversarial Networks (GANs) have been extensively used in image domain transfer tasks. State-of-the-art GAN models have attempted to improve vehicle detection accuracy in night-time by converting infrared images to day-time RGB images. However, these models have been found to under-perform during night-time conditions compared to day-time conditions. Therefore, this study attempts to alleviate this shortcoming by proposing three different approaches based on GAN models that tries to reduce the feature distribution gap between day-time and night-time infrared images to improve vehicle detection accuracy in night-time conditions. Quantitative analysis to compare the performance of the proposed models with the state-of-the-art models have been done by testing the models using state-of-the-art object detection models. Both the quantitative and qualitative analyses have shown that the proposed models outperform the state-of-the-art GAN models for vehicle detection in night-time conditions, showing the efficacy of the proposed models.

Keywords: thermal images, vehicle detection, generative adversarial networks

INTRODUCTION

Object detection algorithms have improved drastically over the past few years, both in terms accuracy of detection and frame rate (running time speed)) Zhao et al. (*1*) and have been a popular area of research in the field of computer vision. Performance of the state-of-the-art object detection algorithms has been observed to improve significantly in day-light conditions using deep learning models. However, the scenario is drastically different under low illumination and poor weather conditions, where traffic cameras are still required to perform well for tasks such as surveillance and advanced driver assistance systems (ADAS). The poor performance of object detection models in night-time conditions can be attributed to the glare from the headlights and tail-lights of vehicles, along with low-illumination which leads to lack of vehicle feature visibility. Consequently, object detection models frequently miss vehicles due to the intense glare or poor illumination. Therefore, there is a significant need for research and innovations in improving vehicle detection performance under low illumination and poor weather conditions.

Use of thermal infrared (TIR) cameras have always been a popular research topic in ADAS to improve night-time driving by enhancing driver's perception. Also, infrared cameras have the advantage that they do not sense the illumination conditions, but the thermal radiations instead. Their robustness to illumination changes and shadow effects makes them useful in ADAS which is the key motivation of this research to use thermal infrared cameras for improving vehicle detection performance under night time conditions. One of the existing popular ways to use infrared images for vehicle detection purpose is to convert them into day-time RGB images, where vehicle detection already performs significantly well. This falls under image-translation task, which involves converting images from one domain (e.g., TIR image) to another (day-time RGB image). With the advancement in deep learning algorithms, a particular class of generative models, namely Generative Adversarial Networks (GANs), have become popular in generating images from random noises and also conditoned data, such as input images. The conditonal GANs (CGANs) Goodfellow et al. (*2*) have been extensively used in image domain translation tasks Isola et al. (*3*) . These models have been successful to synthesize fake images which are realistic and similar to input images domain Goodfellow et al. (*2*). These conditional GANs are capable of image translation from one domain to the other given aligned data for supervised learning approaches (e.g., pix2pix) Isola et al. (*3*) and unaligned data for unsupervised learning based approaches (e.g., cycleGAN) Zhu et al. (*4*) in which generated results are conditioned on input images. Such CGANs have been used in variety of tasks like semantic segmentation, grayscale visible image colourization, sketch to scene etc. Therefore, in this research we have used conditional GANs to convert TIR images to day-time RGB to leverage robustness of infrared cameras to illumination conditions.

Recent studies on converting TIR images to visual RGB conversion has been done using CNN networks by Berg et al. (*5*) and improved further using GAN methodology by Kuang et al. (*6*) in their model, named as Thermal Infrared Colorization using GANS or TIC-GAN. However, these existing GAN models such as cycleGAN, pix2pix, and TIC-GAN have been trained and tested on primarily on day-time images only. This is because the supervised models such as pix2pix or TIC-GAN requires aligned dataset (day-time TIR and RGB images) to generate fake day-time RGB images. On the other hand, unsupervised models such as cycleGAN which doesn't require aligned dataset cannot perform as good as the supervised models. Therefore, existing studies have used only day-time TIR images for image translation tasks. However, there is typically a distinct difference in the features in day-time and night-time TIR images, which can be also observed in Figure 1. Therefore, this leads to difference in training (day-time) and testing (night-time) data distribution, which can lead to poor performance of the GAN models in night-time conditions. This is because the supervised TIC-GAN model is trained using paired day-time RGB and infrared images. Therefore, in this study, we have proposed three models which uses the existing cycleGAN and TIC-GAN models as building blocks, and attempts to reduce the training (day-time images) and test data (night-time images) feature distribution gap, which can help to improve the vehicle detection accuracies. We have tested our proposed approaches on the KAIST dataset (*7*), both qualitatively and also quantitatively using vehicle detection accuracies and compared their performance with the baseline TIC-GAN model.

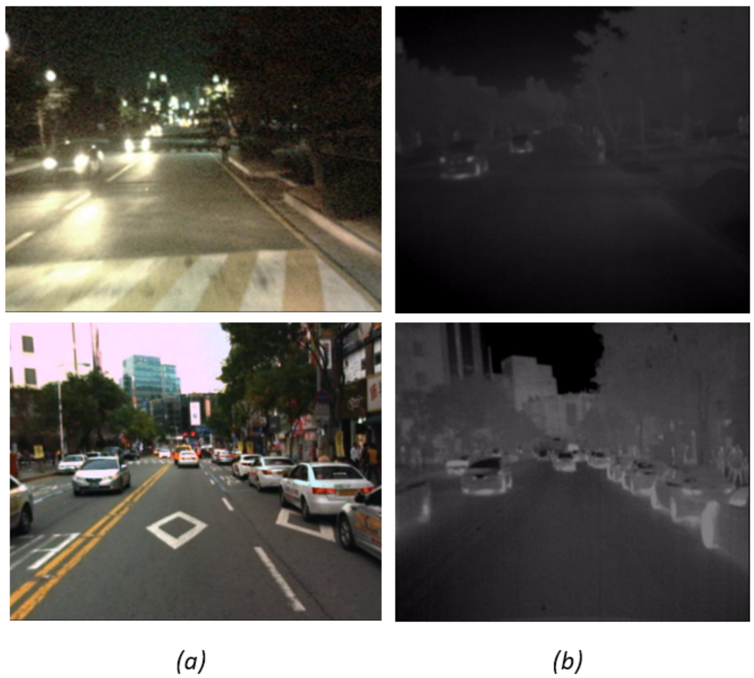


FIGURE 1: Difference in night-time and day-time images (a) RGB Ground truth (b) Infrared Image at night-time and day-time conditions

The following section presents the relevant literature and the motivation behind the study. In the third section implementation of models is done using existing studies and our proposed approaches are presented for the problem statement. The fourth section shows the results obtained using implementation of different models and proposed approaches along with discussion of the results. The fifth section concludes this study by featuring its contributions and identifying future extensions.

LITERATURE REVIEW

In this research, the main focus is to convert TIR images to day-time RGB using Generative adversarial networks (*2*) for improving vehicle detection accuracy in scenario of low illumination conditions and glare coming from vehicles. Quite a few existing studies have been done already based on leveraging infrared images for vehicle detection.

Nam and Nam (*8*) extracted vehicle features from infrared images in low illumination conditions using foreground extraction and classified vehicles based on vehicle-front features like headlights, grills regions, windshield, etc. using bayesian classifier. The method extracts regions of interest (ROI) or vehicular features from the foreground extracted images and texture features of ROI's are quantitatively obtained using Gray-Level Co-occurrence Matrices (GLCMs). Based on the textural features, vehicle classification was done using Bayesian classifier. However, a major drawback of the method in low-illumination condition is the low classification accuracy due to poor spatial resolution of thermal infrared-images to obtain fine texture details.

Ma et al. (*9*) performed thermal infrared and visual grayscale image fusion, to differentiate targets and backgrounds on the basis of difference in radiation. One of the major advantage of the method is that it can work competently in all-season and all-day and night conditions. Although the study was not focused for vehicle detection, however it can be used in the case customized for vehicles as well. However, pixel-wise image registration is required for accurate fusion of infrared and visible images which makes the application difficult. Further, spatial resolution of visible and thermal images is required to be consistent which escalates the difficulty in application. Wang *et al.* (*10*) advanced the flexibility of image enhancement in poor-illumination images. Their study used coloured picture correction primarily on non-linear transformation basis in compliance with the illumination reflection model and multi-scale theory. The primary downside of the study is that it can not be used to improve images based on video and real-time performance of the algorithm is poor.

Recently, GANs have been used extensively for image to image translation tasks and therefore, researchers have also focused on using GANs for converting infra-red images to RGB domain. Pix2pix GAN (*3*), one of the earliest GAN model for image translation tasks, is a conditional GAN model based on supervised learning approach. The supervised data contains images from two domains having perfect pixelwise correspondence. The overall loss function is based on adversarial loss and pixel distance loss which drives the optimization process. However, this model also requires pixelwise correspondence, which is difficult to obtain between RGB and infra-red images. Also feature-wise comparison of generated result and target ground truth is missing in pix2pix which can lead to lack of fidelity in generated results.

CycleGAN (*4*), on the other hand, is also a conditional GAN model based on unsupervised learning. Unlike pix2pix (*3*), it does not not require supervised dataset for image to image translation tasks. This makes it suitable for the cases where paired or aligned dataset is not available *e.g.* zebra to horse conversion, oranges to apple conversion *etc.* Unsupervised image to image translation is a highly unconstrained problem which is solved by CycleGAN using cycle consistency loss. Similar to pix2pix for generating high fidelity output, cycleGAN also uses an adversarial loss. However, even in CycleGAN, lack of feature-wise comparison can make the translated results lack fidelity. Also it is difficult to include feature-wise comparison in cycleGAN unlike pix2pix model. This is because cycleGAN is based on unsupervised learning and we do not have a target output for a particular input image for feature based comparison.

On the other hand, Berg *et al.* (*5*) presented two CNN based approaches to transform thermal infrared to RGB image. Their technique is robust to pixel misalignments between image pairs unlike pix2pix. But the downside of their study is that the generated RGB results are hazy, blurred and lacks fine details. Recently, Thermal Infrared colorization GAN or TIC-GAN (*6*) was proposed, which is a GAN network similar to pix2pix (*3*). However unlike pix2pix, it contains additional loss functions namely, perceptual loss (*11*) and total variational loss (*12*). The perceptual loss uses a pre-trained network based on image classification *e.g.* VGG-16 for extracting feature representations from the image. These feature representations are used for comparison between generated and target ground truth image. The perceptual loss increases fidelity significantly in the generated results. Total variational loss increases spatial smoothness in the generated results by repairing noisy pixels. Apart from the total variational loss and perceptual loss, TIC-GAN model is similar to pix2pix model. For producing high resolution outputs, TIC-GAN uses pix2pix-HD generator network (*13*).In the next chapter, we discuss the details of the proposed approaches for converting night-time TIR to daytime RGB images.

METHODOLOGY

Vehicle detection accuracy on RGB images often drops substantially in combined case of low illumination conditions and glare coming from vehicles during night-time conditions. In such situations, thermal infrared images which only sense thermal radiations can be thought to be used as a feasible solution. One approach can be to convert infrared image to day-time RGB image to overcome the problem of low-illumination and glare from vehicles simultaneously. To implement the conversion of infrared to RGB, GAN can be a suitable choice as it have been extensively used in image domain translation tasks. This section starts with the introduction of state-of-the-art GANs which have been used relevant studies to convert thermal infrared to RGB images. Finally, we introduce and describe in detail about the three different approaches proposed in this study to convert the night-time infrared to day-time RGB images for enhancing vehicle detection accuracy in low-light.

State-of-the-art GAN models for infrared image translation

Goodfellow *et al.* introduced Generative Adversarial Networks (GANs) (*2*) in 2014 which was competent of producing realistic looking fake images by variation of input noise vector to generate changes in output features. The model optimizes itself such that the generated data distribution lies close to the input dataset distribution.

This is done using two networks: the generator network (G) and the discriminator network (D). The generator network G is an encoder-decoder network which learns to upsample a random noise vector z sampled from distribution p_z to a realistic looking fake image $G(z)$. The model aims to bring $G(z)$ closer to real looking samples from the distribution $p_{data}(x)$, where x is the training data. This is achieved with the help of the discriminator network D which is a classifier network that tries to distinguish between the real samples x and fake samples $G(z)$ and scores the generator on the basis of realness of the output $G(z)$ it produces. This score is called the generative adversarial loss. The generator network tries to maximize this loss whereas the discriminator tries to minimize it. This *minimax* game ultimately drives the generator network to produce high fidelity fake images $G(z)$ and the generated data distribution p_g eventually comes close to training data distribution p_{data}.

The adversarial loss function is expressed as shown below :

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{1}$$

However, the output of basic GAN network (*2*) has a major drawback that it cannot generate desired class of output as the noise vector is random. This problem can be addressed using conditional GANs. Conditional GAN (*14*) made controllable data generation possible. The generator and discriminator networks in conditional GANs are trained on some additional auxiliary data *e.g.* class labels or information obtained using other modes. We discuss next the details of the cycleGAN and TIC-GAN models, which form the building blocks of the approaches proposed in this study.

CycleGAN

CycleGAN, introduced by Zhu *et al.* in 2017 (*4*), is useful for image domain translation tasks where supervised dataset is not available. Unlike supervised methodology based GAN models, CycleGAN has an advantage that it does not require a target output and it learns about the style and content from the target domain to generate fake images in the target domain. The problem of unsupervised image domain transfer is highly unconstrained and CycleGAN handles this using image reconstruction by including cycle consistency loss. CycleGAN includes an adversarial loss for generating realness in the generated outputs. For further improvement in quality of output, CycleGAN consists of two generator discriminator network pairs in series to put an additional constraint of mapping in both forward and backward directions. Both of the generator networks and both the discriminator networks are exactly similar in architecture. The discriminator network is the PatchGAN discriminator network. PatchGAN, focus on penalizing image patches separately rather than penalizing complete image at once. To see the complete explanation about Cycle GAN loss function and architecture details please refer to the CycleGAN paper (*4*). In this study we have used cycleGAN as one of the components of the pipeline to improve night-time infrared to day-time conversions and its application can be seen in the proposed approaches explained in this study.

Thermal Infrared colorization using Conditional GAN

The Thermal Infrared colorization via Conditional GAN or the TIC-GAN was introduced in 2017 by Kuang *et al.* (*6*). It was found to perform better than the previous studies or baselines for infrared to visual RGB conversion task *eg*. infrared colorization by Berg *et al.* (*5*), Pix2pix GAN (*3*) etc. Infrared colorization has pixelwise misalignments present due to sensor differences in thermal and RGB cameras, and this is one of the reasons why previous studies have shown poor results in the task. TIC-GAN leverages supervised data for the training process. However it produces qualitatively better results for the infrared colorization problem by including feature based comparison rather than only focusing on pixel to pixel comparison between target and generated images. To address this unique problem of infrared colorization, TIC-GAN uses feature based loss function called as perceptual loss (*11*). This helps model in developing feature based understanding of the domains and the generated images can be expected be more robust to small pixelwise changes. Also, it includes total variational loss which results in significant noise reduction in generated output. TIC-GAN has a generator network to perform infrared to visual RGB domain translation and a discriminator network to optimize the generator network's performance. The TIC-GAN model uses an additional VGG-16 network for feature based comparison of generated and target images *i.e.* for perceptual loss. The model uses the pix2pix HD (*13*) generator network given by Zhu *et al.* in 2018 to obtain high resolution generated images of superior quality. The discriminator network in TIC-GAN is similar to the cycleGAN model *i.e.* the PatchGAN discriminator. For more details on loss function, network architecture, please refer to the TIC GAN paper (*6*). The TIC-GAN model is the building block of the proposed approaches in this study. Also, it has been chosen as the baseline for comparison of results.

Proposed approaches for the problem

Converting infrared image to its corresponding day-time RGB is an interesting task, because unlike grayscale image to RGB conversion which requires only estimation of chrominance *i.e.* color, infrared to RGB conversion requires estimation of both chrominance and luminance. Popular conditional GAN models like PIX2PIX (*3*)(see Figure 20 of (*3*)) and CycleGAN (*4*) perform poorly in the above mentioned task since getting both unknown entities *i.e.* chrominance and luminance may also require feature based understanding. The TIC-GAN model (*6*), on the other hand, holds the capability of converting day-time infrared input to day-time RGB images. However, the model was not found to perform well in converting night-time infrared images to their corresponding day-time RGB images. The primary reason behind this being the gap in training input (day-time infrared) and testing input (night-time infrared) image feature distribution.

Major difference in feature distributions of infrared images at night-time and day-time are the contrast conditions. The contrast conditions in the infrared images can be found to be better in the day-time than the night-time infrared images (see Section 4.6 of (*5*)). In simple terms the surroundings at night adopts a more homogeneous temperature, thus making contrast in the thermal infrared images in the night to be lower than the day-time infrared images. Reducing this training and testing feature distribution gap is therefore required to be done, in order to keep the test performance at par with the train performance. To achieve this, we have proposed three different models that use the advantages of the existing GAN models such as CycleGAN (*4*) and ToDay GAN (*15*). The three different proposed approaches are presented in the subsequent subsections that attempts to reduce the gap in training and testing data distribution to improve test results both qualitatively and quantitatively. Training details of proposed approaches are explained in the "Dataset and Training Details" subsection at the "Results and Discussion" section.

Proposed approach-1

In this approach, efforts have been made to reduce the training and testing input distribution gap by making testing data input similar to the training data input. Keeping this in mind, the test input which is the night-time Infrared can be converted to day-time infrared domain. Because TIC-GAN model performs satisfactorily on day-time infrared input, if the conversion of night-time Infrared to day-time infrared is realistic enough *i.e.* on modifying the test input to be similar to training input, test results are anticipated to improve.

Since the training of the TIC-GAN model is done on day-time infrared and corresponding day-time RGB paired dataset, to generate day-time infrared images from night-time infrared images, CycleGAN can be used for this unsupervised domain translation task. Finally fake day-time infrared images obtained by converting night-time test images using CycleGAN can be used as an input to TIC-GAN.

The pipeline for the network is shown in the Figure 2 which shows that query night-time-image is first converted to day-time infrared domain image using CycleGAN. This fake day-time infrared image is served as an input to the TIC-GAN model which is trained on day-time infrared and corresponding day-time RGB aligned dataset. The TIC-GAN model when given the fake day-time infrared input finally generates the fake day-time RGB image.

Proposed approach-2

This approach keeps training and testing input to be from the same domain i.e. night-time infrared unlike proposed approach-1 in which test input was modified to be consistent with training input. For converting night-time infrared to day-time RGB this approach uses ToDay GAN (*15*) as one of the components along with TIC-GAN. ToDay GAN (*15*) is based on unsupervised domain translation for night-time RGB images to day-time RGB. To reduce the feature gap between training and testing input as mentioned in the starting of this section ("Proposed approaches for the problem"),

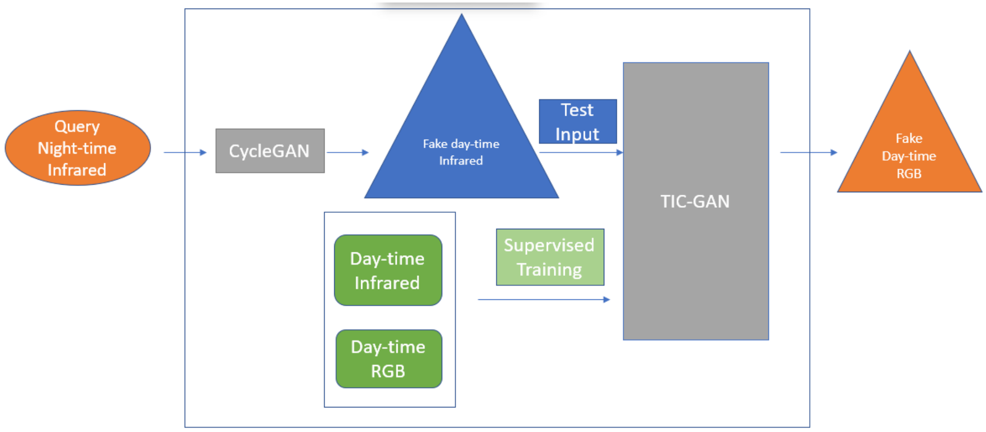


FIGURE 2: Pipeline for the proposed approach-1

it can be a better idea to train and test on the same kind of dataset *i.e.* the training and testing should both be done on night-time infrared images. Since we want to convert night-time infrared input to day-time RGB and we want to use night-time infrared images in the training dataset, we do not have a corresponding day-time RGB ground truth. To alleviate that, we have leveraged ToDay GAN to convert the night-time RGB to fake day-time RGB ground truth. Ultimately, this can provide night-time infrared images for both training and testing, and the TIC-GAN model can be trained with night-time infrared and corresponding day-time RGB obtained using ToDay GAN.

Using this, we can synthesize supervised dataset consisting of night-time infrared and corresponding synthesized fake day-time RGB. Therefore, the TIC-GAN model can be trained directly on night-time infrared and corresponding day-time RGB, testing can be done on night-time infrared. The pipeline of the proposed approach is shown by Figure 3.

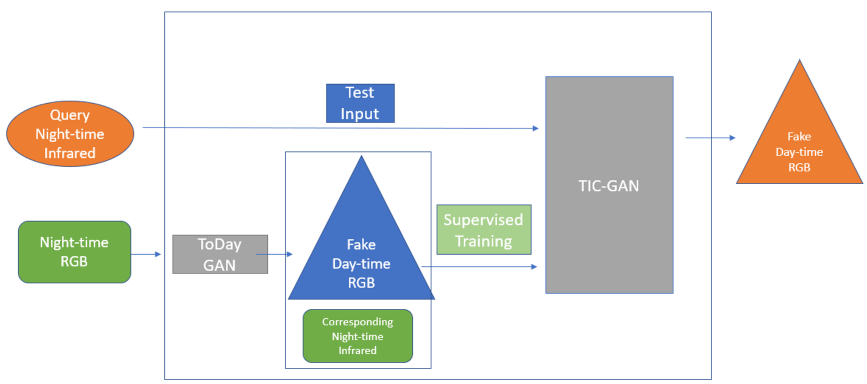


FIGURE 3: Pipeline for the proposed approach-2

Proposed approach-3

In this approach, to bring the training and testing data distribution as close as possible as mentioned in starting of this section ("Proposed approaches for the problem"), we can modify training data input (day-time infrared image) to look similar to testing data input domain (night-time infrared image). To convert day-time infrared to fake night-time infrared, CycleGAN(unsupervised domain translation) have been used. The resulting fake night-time infrared is paired with the corresponding ground truth day RGB. Training of TIC-GAN model is then done on synthesized night-time infrared (synthesized from day-time infrared input using CycleGAN) and corresponding ground truth day-time RGB, and finally we test the TIC-GAN model on the night-time infrared query image. This can help to bring the feature distributions of infrared images closer to each other during training and testing.

The idea is to bring query day-time infrared images as close as possible to night-time infrared. The better this conversion would be, the more realistic results we will be getting. CycleGAN model is used for the conversion of day-time infrared to night-time infrared conversion. The higher the quality of conversion of day-time infrared to night-time infrared, better will be the quality of results that will be obtained at the test time from input night-time infrared using the TIC-GAN model. Figure 4 shows the pipeline for the proposed approach-3 for better visualization of the approach. In the next section, we discuss in details about the results obtained from various proposed approaches and the baseline TIC-GAN model, their shortcomings, qualitative and quantitative evaluation of the proposed approaches and TIC-GAN.

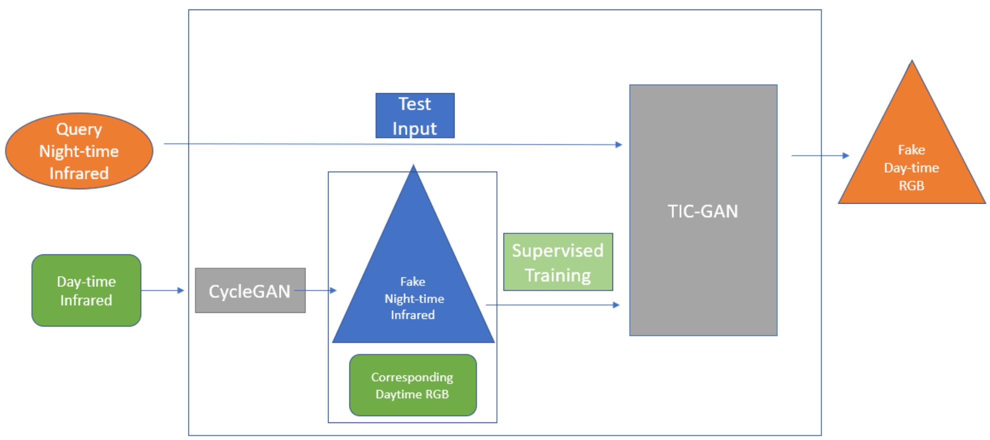


FIGURE 4: Pipeline for the proposed approach-3

RESULTS AND DISCUSSIONS

In this section, we provide the detailed results obtained from the three proposed approaches in this study along with the comparison with the TIC-GAN model as the baseline algorithm. We first discuss the details of the dataset on which the algorithms have been trained and tested, along with the training details of each model. Then we show the qualitative comparison of the models, followed by the quantitative comparison using object detection performance metrics.

Dataset and Training Details

All experiments in this study has been performed using the KAIST thermal dataset (*7*). This dataset consists of 95k thermal-RGB pairs (640x480, 20Hz) taken from a vehicle. Out of this, about 66% of the image pairs are of day-time and rest are night-time images.

In proposed approach-1, cycleGAN model is trained to transform query night-time infrared image to day-time infrared. And for doing so, it is trained on the night-time infrared and day-time infrared domains. We used 5000 images each from both day-time and night-time infrared domains. Model is trained for 60 epochs, with learning rate of 0.0002 for first 40 epochs which is further decayed to 0 over the next 20 epochs. Rest all the hyper-parameters are set to their default values as mentioned in (*4*). Also, TIC-GAN model is used in proposed approach-1 for converting infrared to RGB, and it is trained on KAIST paired thermal-RGB dataset with 5000 image pairs. The model is trained for 100 epochs, 80 epochs with a learning rate α=0.0002 which is gradually reduced to 0 in the next 20 epochs. Other aspects and hyperparameters while training are exactly similar to the values used in the original TIC-GAN paper (*6*).

In proposed approach-2, for training ToDay GAN model KAIST dataset is used in unaligned manner with good illumination night-time images in one domain and day-time images in the other domain. Training was done for 80 epochs. Learning rates are started at 2e-4 for generator networks and 1e-4 for discriminator networks, and are kept constant for the first training half and linearly decreased to zero in the second half of training. λ is the hyperparameter for cycle consistency loss is set to 10.0. Also, TIC-GAN model is used in proposed approach-2 for converting night-time infrared to day-time RGB. For doing so, corresponding day-time RGB for night-time infrared is obtained by using ToDay GAN by converting night-time RGB to day-time RGB. This synthesized dataset having night-time infrared and synthesized day-time RGB consists of 600 image pairs. The model is trained for 100 epochs, 80 epochs with a learning rate α=0.0002 which is gradually reduced to 0 in the next 20 epochs. Other aspects and hyperparameters while training were exactly similar to the values used in the original TIC-GAN paper (*6*).

In proposed approach-3, CycleGAN is trained to transform day-time infrared to night-time infrared. For which it is trained on the day-time infrared and night-time infrared domains. We used 5000 images each from both day-time and night-time infrared domains. Rest training settings for Cycle-GAN are exactly similar to the cycleGAN model used in proposed approach-1. TIC-GAN model is trained to transform night-time infrared to day-time RGB. For training TIC-GAN, we use fake night-time infrared obtained from cycleGAN model which is paired with day-time RGB. Synthesized dataset contains 5000 image pairs. Other training settings for TIC-GAN are exactly similar to proposed approach-1 TIC GAN model.

Qualitative Analysis

In this study, three different approaches have been proposed to reduce the feature distribution gap between the night-time TIR and day-time TIR while training and testing to improve night-time TIR to day-time RGB conversion. Figure 5 shows the sample results of the images obtained from the baseline TIC-GAN, proposed approach-1, proposed approach-2, and proposed approach-3, and the ground truth RGB images. It can be seen from the figure that the proposed approach-3 performs better than the TIC-GAN and the other proposed approaches. Significant blurring is present in the outputs from TIC-GAN and proposed approach 1 and 2, which can make the vehicle detection difficult. Further, some artifacts such as pink patches were also observed in outputs from the proposed approach 2, as shown in Figure 4d, top row. Please note that the baseline TIC-GAN in this study has been tested on night-time images, unlike the study by Kuang *et al.* (*6*), where TIC-GAN was proposed and tested on day-time images. Since the main focus of this study is to improve vehicle detection in night-time conditions, testing on the night-time infrared images shows that the shortcomings of TIC-GAN under such low illumination conditions, where proposed approaches, in particular proposed approach 3 perform significantly better.

In addition to overall qualitative evaluation of the quality of the generated images, we also need to evaluate the models' performance with respect to vehicle detection. Vehicle detection using object detection algorithms *e.g.* YOLOv5 (*16*) is reliably accurate in day-time conditions.

FIGURE 5: Qualitative comparison of the TIC-GAN and proposed approaches : (a) Night-time Infrared (b) TIC-GAN (c) Proposed approach-1 (d) Proposed approach-2 (e) Proposed approach-3 (f) Ground Truth RGB

However it is vulnerable to low accuracy in low-light conditions due to noisy images, insufficient illumination to reveal vehicle features and also due to glare from vehicle headlights and tail-lights towards the camera.

Figure 6 shows the comparison of object detection results from the groundtruth night-time RGB, TIC-GAN, and the proposed approaches. The top two rows show images from good illumination conditions due to street lights, while the bottom two rows are from images under poor illumination conditions. It is evident from the figure that the low-light environment is challenging for vehicle detection and vehicles are highly vulnerable to be omitted for detection (false negatives), due to combined effect of glare from vehicles and low illumination conditions which make vehicle detection omission more likely. In contrast, the proposed approach-3 shows considerable reduction in such false negatives *i.e.* lesser omission of vehicles due to low-illumination and glare. Further, the artifacts such as pink patches observed in outputs from proposed approach 2 also found to not impact the vehicle detection performance.

Under good illumination conditions (Figure 6a top two rows), we can observe that object detection in case of night-time RGB have lesser false negatives (vehicle omission in detection). This can be attributed to the fact that even in case of glares from vehicle headlights, the vehicle features are still sufficiently visible due to good illumination conditions. Also, conversion from night-time infrared to day-time RGB often produces noisy patches in outputs which can in turn increase the vulnerability for higher false positive rate (false alarm for object detection) than that of object detection on good illumination night-time images. Nonetheless, it can be observed from the Figure 6 that proposed approach-3 has lower false positives (false vehicle detection) and lower false negatives (vehicle detection omission) as compared to TIC-GAN, proposed approach-1, proposed approach-2 under poor illumination conditions. Next, we discuss the quantitative comparison of the proposed approaches using veicle detection accuracies.

FIGURE 6: Object detection results comparison: (a) Groundtruth Night-time RGB (b) TIC-GAN (c) Proposed approach-1 (d) Proposed approach-2 (e) Proposed approach-3

Quantitative Evaluation

From the results described above, it can be observed that the proposed approach-3 works best in case of night-time TIR to day-time RGB translation task qualitatively, particularly under low illumination night -time conditions. Quantitatively, this is not an easy task for comparison since groundtruth day-time RGB images are not available for comparison with the generated day-time images, which can be used for pixelwise comparison. Hence, in this study, we have used object detection metrics as a quantitative comparison criteria instead. We have used the state-of-the-art You Only Look Once (YOLOv5) model (*16*), trained on COCO dataset for object detection on the generated images. These test metrics can serve as a quantitative comparison criteria of results generated using different approaches.

All vehicles (cars, buses, trucks, motorcycles, and bicycles) were manually annotated in 200 ground truth night-time RGB images to create the test dataset for vehicle detection comparison. These annotated vehicles were then compared with the same vehicle classes detected using YOLOv5 on the ground-truth RGB images and the images generated using TIC-GAN and the proposed three approaches. Further, a separate test dataset containing 200 day-time images has been annotated and tested directly using YOLOv5, which can be used as the performance criteria, that the proposed approaches should ideally achieve. The groundtruth test RGB images are sampled randomly from the KAIST dataset, used in this study.

Table 1 shows the precision, recall, and mAP@0.5 on the generated results obtained using the three different proposed approaches, TIC-GAN, corresponding ground-truth night-time RGB, and the day-time RGB for the test dataset. The number of images on which each model is tested along with the number of vehicles that were present in the images and annotated are also mentioned in the Table 1. Further, the precision-recall (PR) curve, obtained from this test dataset, is shown in Figure 7. Here, *AP* or the average precision is the area under PR-curve and mean of *AP*'s for all categories of vehicles on which the model is trained on to detect gives the mAP. The mean average precision values are obtained at IoU threshold of 0.5 also called as "mAP@0.5".

From the Figure 7, it can be seen that the PR-Curve of the three proposed approaches outperforms the state-of-the-art TIC-GAN model. This is also observed from the precision, recall, and *mAP*@0.5 values shown in Table 1. In particular, proposed approach 3 performs better compared to the TIC-GAN and the remaining two proposed approaches, which was also observed based on the qualitative evaluation results.

TABLE 1 Object detection evaluation results

Model	Images	Annotated Vehicles	Precision	Recall	mAP@0.5
TIC-GAN	200	987	0.563	0.342	0.347
Proposed approach-1	200	987	0.658	0.348	0.441
Proposed approach-2	200	987	0.734	0.480	0.571
Proposed approach-3	200	987	0.681	0.614	0.667
Night-time RGB	200	987	0.861	0.653	0.797
Day-time RGB	200	1247	0.877	0.881	0.907

However, *mAP*@0.5 and the PR-curve obtained from the groundtruth night-time RGB images are still found to be better than the baseline and all the proposed approaches. One of the possible reason behind this observation can be that the majority of night-time images in KAIST dataset (*6*) contains good illumination night-time images. On the other hand, most of the low illumination images in KAIST dataset had empty streets and very few vehicles were present in such images which made the test dataset and consequently the analysis more biased towards the good illumination night-time images. Figure 8 shows some sample low illumination images with empty streets, along with the output of fake day-time image obtained using proposed approach 3. Further, the bottom two rows of images shown in Figure 6 shows the performance of different models under low illumination conditions, where the proposed approach-3 can be seen to perform better than the baseline TIC-GAN model.

Therefore, it can stated that the proposed approach-3 has shown good results in night-time infrared to day-time RGB conversion task and its use is specifically important for perceiving night-time environment in case of low illumination conditions. It can be seen that the model is robust towards poor illumination conditions and we can leverage this property for better perception of dark environment in advanced driving assistance systems and surveillance applications.

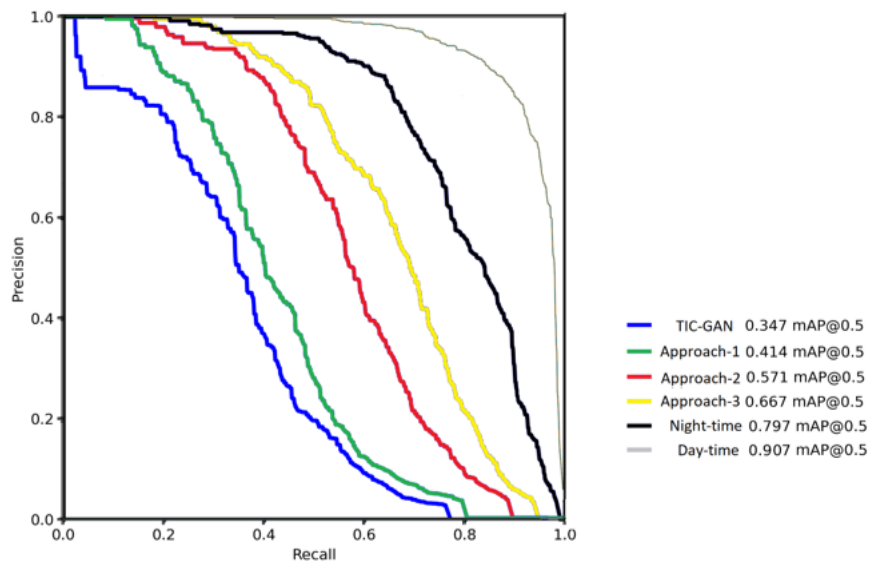
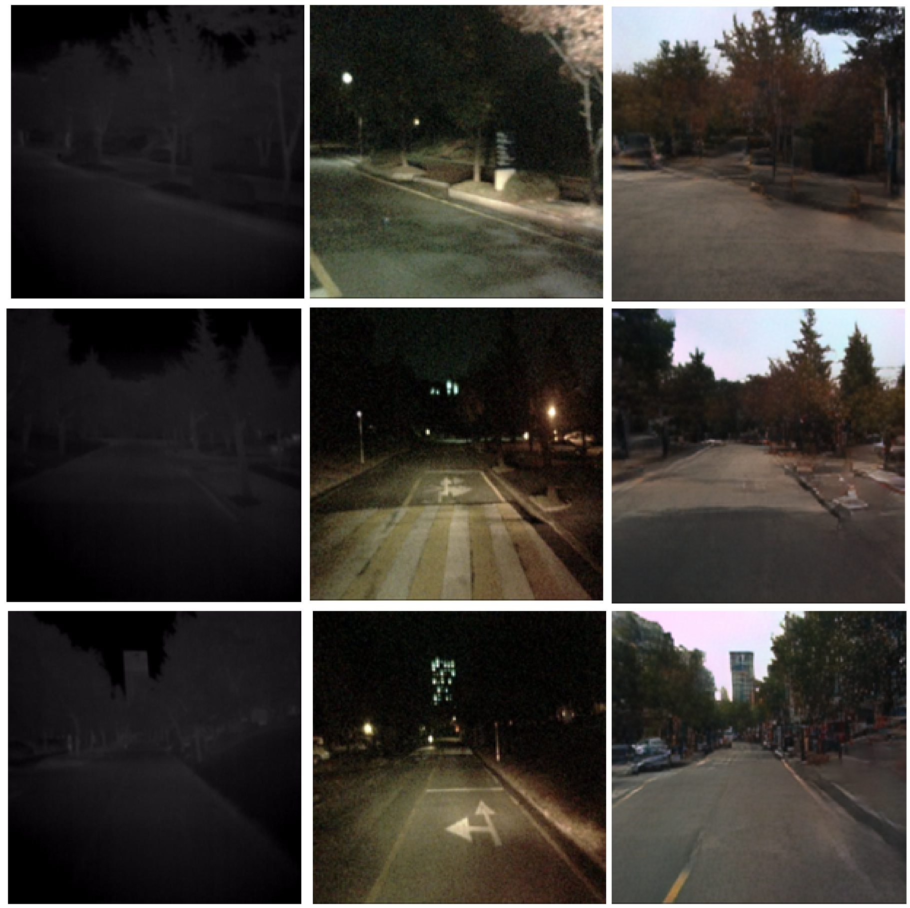


FIGURE 7: Precision-Recall curve on object detection results

CONCLUSION

In this research, we have proposed three different approaches to handle low accuracy of vehicle detection in dark environments or poor-illumination conditions using thermal infrared images. Glare from vehicles and low-illumination conditions combinedly results in poor vehicle detection performance. However, infrared images are robust to illumination changes and dependent on thermal conditions of the environment. Keeping this in mind, to address this problem of improving vehicle detection accuracy in low-light conditions, we have converted night-time infrared images to day-time RGB images using GAN networks. GAN models such as TIC-GAN, which has been used in literature for conversion of infrared to RGB images, have been found to perform well in day-time images, but their performance drops significantly in night-time conditions. This is because the supervised TIC-GAN model is trained using paired day-time RGB and infrared images. Therefore, in this study, we have proposed models which uses the existing cycleGAN and TIC-GAN models as building blocks, and attempts to reduce the training (day-time images) and test data (night-time images) feature distribution gap, which can help to improve the vehicle detection accuracies.

The proposed approach-1 first converts the query night-time infrared image to day-time infrared image using cycleGAN which forms the input to the TIC-GAN model to output fake day-time RGB image. On the other hand, proposed approach-2 converts night-time RGB to fake day-time RGB using ToDayGAN, which is used to train the TIC-GAN model using fake daytime RGB and corresponding night-time infrared image. The training data of TIC-GAN model in the proposed approach-3, however consist of fake night-time infrared (obtained using cycleGAN) and corresponding day-time RGB image. These attempts help to reduce the training and test data distribution gap. We test our proposed approaches on the KAIST dataset (*7*), both qualitatively and quantitatively using vehicle detection accuracies and compare their performance with the baseline



(a) (b) (c)

FIGURE 8: Dark environment perception in night-time : (a) Night-time Infrared (b) Ground truth RGB (b) Output from Proposed approach-3

TIC-GAN model. Our analyses show that the proposed approach-3 outperforms the baseline and the other proposed approaches. Significantly less artifacts and blurring is observed in the outputs obtained using the proposed approach-3 and the model performs well even in the challenging low-illumination night-time conditions.

However, the object detection evaluation metrics of night-time RGB has been found to be higher than the proposed approach-3 based detection. This can be due to the sparse representation of poor illumination based night-time traffic images in the KAIST dataset (*7*)), which also has mostly empty streets with very few vehicles. Due to this fact, majority of analysis is more biased towards good illumination condition based night-time traffic scenes, where existing object detection models already performs fairly well. Thus, it can be concluded that while object detection performs fairly well in good illumination night-time conditions, however their performance drops significantly under low-time illumination conditions where the proposed approach-3 can be used based on thermal infrared images to improve vehicle detection accuracies. In future, the proposed approaches can be trained and tested on large-scale self-curated challenging low illumination images to further determine the efficacy of the models. Further, models based on fusion of RGB and thermal infrared images, both in training and inferencing time, can also be looked upon to improve the vehicle detection accuracies under all illumination conditions.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design: S. Bhargava, P. Chakraborty; data collection: S. Bhargava; analysis and interpretation of results: S. Bhargava, P. Chakraborty; draft manuscript preparation: S. Bhargava, P. Chakraborty. All authors reviewed the results and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

Our research results are based upon work supported by the Initiation Grant scheme of Indian Institute of Technology Kanpur (IITK /CE/2019378). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the IITK.

REFERENCES

1. Zhao, Z.-Q., P. Zheng, S.-T. Xu, and X. Wu, Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 11, 2019, pp. 3212–3232.
2. Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks. *Communications of the ACM*, Vol. 63, No. 11, 2020, p. 139–144.
3. Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros, Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
4. Zhu, J.-Y., T. Park, P. Isola, and A. A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.
5. Berg, A., J. Ahlberg, and M. Felsberg, Generating Visible Spectrum Images from Thermal Infrared. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.
6. Kuang, X., J. Zhu, X. Sui, Y. Liu, C. Liu, Q. Chen, and G. Gu, Thermal infrared colorization via conditional generative adversarial network. *Infrared Physics & Technology*, Vol. 107, 2020, p. 103338.
7. Hwang, S., J. Park, N. Kim, Y. Choi, and I. S. Kweon, Multispectral pedestrian detection: Benchmark dataset and baseline. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.
8. Nam, Y. and Y.-C. Nam, Vehicle classification based on images from visible light and thermal cameras. *EURASIP Journal on Image and Video Processing*, Vol. 2018, No. 1, 2018.
9. Ma, J., Y. Ma, and C. Li, Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, Vol. 45, 2019, p. 153–178.
10. Wang, W., Z. Chen, X. Yuan, and X. Wu, Adaptive image enhancement method for correcting low-illumination images. *Information Sciences*, Vol. 496, 2019, p. 25–41.
11. Johnson, J., A. Alahi, and L. Fei-Fei, Perceptual Losses for Real-Time Style Transfer and

Super-Resolution. *Computer Vision – ECCV 2016 Lecture Notes in Computer Science*, 2016, p. 694–711.

12. Chen, Q., P. Montesinos, Q. S. Sun, P. A. Heng, and D. S. Xia, Adaptive total variation denoising based on difference curvature. *Image and Vision Computing*, Vol. 28, No. 3, 2010, p. 298–306.
13. Wang, T.-C., M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
14. Mirza, M. and S. Osindero, *Conditional Generative Adversarial Nets*, 2014.
15. Anoosheh, A., T. Sattler, R. Timofte, M. Pollefeys, and L. V. Gool, Night-to-Day Image Translation for Retrieval-based Localization. *2019 International Conference on Robotics and Automation (ICRA)*, 2019.
16. Redmon, J., S. Divvala, R. Girshick, and A. Farhadi, You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.