Depthwise Convolution For Compact Object Detector In Night-time Images (*Supplementary Material*)

I. Ablation Study

In the supplementary material, we show the effectiveness of different modules employed in the proposed object detection module. It includes the role of Adaptive Histogram Equalization (AHE), activation functions (i.e., ReLU, Leaky ReLU (LReLU), Parametric ReLU (PReLU) and Exponential Linear Unit (ELU)), importance of proposed loss function, optimizer and depthwise convolution in the architectural design. We therefore conducted many experiments in ablation study and justify the proposed design by comparing their performance quantitatively with recall rate, f_1 score and mean Average Precision (mAP) which are depicted in Table I.

TABLE I: Results obtained on ablation study on different functions used in the proposed object detection model on FLIR validation dataset. Here, the metrics measurement are carried out by averaging values of three classes (i.e., person, car and bicycle). The bold font texts indicate better values compared to other approaches.

Metrics	**Recall**	f_1 **score**	**mAP**
Adaptive Histogram Equalization (AHE)			
Without AHE	29.86	33.08	39.23
With AHE (Proposed)	**35.54**	**41.80**	**48.36**
Activation Functions			
ReLU	34.89	39.55	46.98
LReLU	35.45	40.87	47.05
PReLU	33.12	40.60	44.43
ELU (Proposed)	**35.54**	**41.80**	**48.36**
Loss Functions			
BCE	35.08	39.28	47.43
BCE_Dice	34.98	40.64	48.22
Focal	34.46	39.99	46.63
IoU	34.18	39.05	48.26
Tversky	35.14	41.04	48.18
BCE_Dice_IoU	34.65	40.19	47.27
Focal_IoU	33.57	39.32	47.65
Tversky_IoU (Proposed)	**35.54**	**41.80**	**48.36**
Depthwise Convolution (DC)			
Without DC	34.64	39.36	46.36
With DC (Proposed)	**35.54**	**41.80**	**48.36**
Optimizer			
SGD	30.86	33.82	40.68
Adam (Proposed)	**35.54**	**41.80**	**48.36**

We show the effectiveness of different modules employed in the proposed object detection module in this section. It includes the role of activation functions (i.e., ReLU, Leaky ReLU (LReLU), Parametric ReLU (PReLU) and Exponential Linear Unit (ELU)), importance of proposed loss function, optimizer and depthwise convolution in the architectural design. We therefore conducted many experiments in ablation study and justify the proposed design by comparing their performance quantitatively with recall rate, f_1 score and mean Average Precision (mAP) which are depicted in Table I.

Effectiveness of AHE

To enhance the details present in the thermal image, initially we pass it through Adaptive Histogram Equalization (AHE). It enhances the tiny details with relatively better contrast. Hence, it is helpful to improve the visibility of night-time images [8], [5], [7]. The proposed method has been experimented with/without employing AHE module that displays in Table I and obtained noticeable results with help of AHE.

Effectiveness of ELU activation function

In deep learning, a network without activation function works as a linear regression model which cannot perform the given task in an effective manner. Hence, it requires a non-linear activation function in order to learn the complicated and complex form of data. During early developing stage of deep learning models, *sigmoid* function was used as an activation function; however, due to the limitation of vanishing gradient problem, a Rectified Linear Unit (ReLU) was chosen instead [6]. It is defined mathematically as, $f(x) = max(0, x)$, where x represents the value of that particular node. Thus, it activates above zero value; hence, its partial derivative is one. Therefore, the problem of vanishing gradient does not exist.

In the proposed method, we replace all activation functions with ReLU and train the network. The mAP score on ReLU can be observed from the Table I and note a lower mAP due to the disadvantage of ReLU activation function which has zero gradient whenever unit is inactive. As a consequence the algorithm would never adjust the weights for initially inactive nodes. In order to overcome such shortcoming, the Leaky ReLU (LReLU) activation function allows small non-zero gradient when nodes are inactive and sacrifices the sparsity for the gradient during optimization. Thus, it eliminates the saturation problem of ReLU [4]. The proposed method has been trained using Leaky ReLU activation function and the result obtained using this experiment is displayed in same the table. One can note an improvement in mAP than that of obtained using ReLU. Further, we train the proposed model with Parametric ReLU (PReLU) [1] which makes the use of coefficient of leakage into a parameter that is learned along with the other neural network parameters. Hence, it improves the model with low computational cost and reduces the risk of over-fitting. However, it reduces the mAP on the proposed model. Additionally, the proposed method is trained on ELU function that tends to converge to zero in minimum time and also produces more accurate results. In comparison to other other activation functions, ELU has an extra α constant which must be positive number [10] and assists to reach optimum convergence. The effectiveness of ELU activation function in the proposed object detection method is observed and hence, we use it in the proposed method.

Effectiveness of the proposed loss function

Loss functions define how neural network models calculate the overall error from their residuals for each epoch. The importance of different loss functions such as Binary Cross Entropy (BCE) loss [15], BCE_Dice loss [12], focal loss [3], tversky loss [9] and IoU loss [13] can be observed from Table I. The BCE is defined as a measure of the difference between two probability distributions for a given random variable or set of events [15]. It is widely used for classification and pixel-level segmentation. The BCE_dice loss is a weighted sum of dice loss [11] and BCE. It attempts to leverage the flexibility of dice loss of class imbalance and at the same time it uses cross-entropy for curve smoothing. Focal loss down-weights the contribution of easy examples and enables the model to focus more on learning hard examples [3]. It works well for highly imbalanced class scenarios. Further, as mentioned earlier Tversky loss [9] adds a weight to False Positives (FPs) and False Negatives (FNs) with the help of constant coefficient and calculates the similarity between two objects. The IoU loss [13] calculates the ratio between the overlapping regions of the positive instances between two sets of objects, and their mutual combined values. Additionally, we train the network with combination of many losses such as BCE_Dice, BCE_Dice_IoU and BCE_Tversky_IoU. The effectiveness of the loss functions utilized in the proposed method (i.e., combination of Tversky and IoU loss) can be verified from Table I.

Effectiveness of Depthwise convolution & Optimizer

Depthwise convolution is a spatial convolution where each channel of input feature maps are convolved with kernel independently instead of processing it on entire volume (i.e., on all channels) [14] requiring lesser number of parameters to adjust. The role of Depthwise Convolution (DC) in the proposed method can also be verified by looking at the results obtained on the proposed method with/without using it as depicted in Table I. In addition, a proper optimizer improves the detection results with better efficiency. Therefore, Adam optimizer is used to optimize the model for training. It takes advantage of adaptive gradients and Root Mean Square (RMS) propagation [2]. The ablation study on Adam optimiser is also carried out by training the proposed method on the Stochastic Gradient Descent (SGD) and depicted the detection scores in the Table I.

REFERENCES

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. pages 1026–1034, 2015.

[2] Diederik P Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *Proc. 3rd Int. Conf. Learn. Representations*, pages 1–15, 2014.

[3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. pages 2980–2988, 2017.

[4] Andrew L Maas, Awni Y Hannun, and Andrew Y Ng. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3, 2013.

[5] Shivangi Mishra, Ashish Dwivedi, and Badal Soni. Foggy image enhancement using improved histogram equalization and guided filter. In *Smart and Intelligent Systems*, pages 443–452. Springer, 2022.

[6] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.

[7] Monal Patel, Arvind Yadav, and Carlos Valderrama. Image enhancement and object recognition for night vision traffic surveillance. In *Soft Computing for Security Applications*, pages 733–748. Springer, 2022.

[8] Stephen M Pizer, E Philip Amburn, John D Austin, et al. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

[9] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. pages 379–387. Springer, 2017.

[10] Anish Shah, Eashan Kadam, Hena Shah, Sameer Shinde, and Sandip Shingade. Deep residual networks with exponential linear unit. In *Proceedings of the Third International Symposium on Computer Vision and the Internet*, pages 59–65, 2016.

[11] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. pages 240–248. Springer, 2017.

[12] Saeid Asgari Taghanaki, Yefeng Zheng, S Kevin Zhou, Bogdan Georgescu, Puneet Sharma, Daguang Xu, Dorin Comaniciu, and Ghassan Hamarneh. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Computerized Medical Imaging and Graphics*, 75:24–33, 2019.

[13] Floris van Beers, Arvid Lindström, Emmanuel Okafor, and Marco A Wiering. Deep neural networks with intersection over union loss for binary image segmentation. pages 438–445, 2019.

[14] Rongtian Ye, Fangyu Liu, and Liqiang Zhang. 3d depthwise convolution: Reducing model parameters in 3d vision tasks. *arXiv preprint arXiv:1808.01556*, 2018.

[15] Ma Yi-de, Liu Qing, and Qian Zhi-Bai. Automated image segmentation using improved pcnn model based on cross-entropy. pages 743–746. IEEE, 2004.