# Literature Review

This paper introduces a single network, object detection algorithm, You Only Look Once. As the name suggests, YOLO frames the bounding boxes by assigning each object a confidence score and classify the object on the basis of how confident it is about the categorization of the object. The paper compares the performance of YOLO to other object detection algorithms like DPM and R-CNN. As a result, YOLO outperforms both the algorithm as it is a one-time evaluation, although it has its own drawbacks[1].

This paper talks about the YOLOv3, YOLOv3 is used which predicts bounding boxes using dimension clustering. Anchor boxes are used in dimension clustering, eventually predicting 4 coordinates for each bounding box. It predicts an object score using logistic regression and uses the sum of squared error loss. If the ground truth and the object score are the same as any other prior bounding box, then the sum of squared error loss is 1. However, If the bounding box priories not the best but does overlap a ground truth object by more than some threshold it will ignore the prediction. The darknet-53 network is used to extract features. This new network is a hybrid approach between the network used in YOLOv2, Darknet-19, and that newfangled residual network stuff[2].

This paper provides a detailed study of the multi-spectral KAIST Dataset. It comprises of a wide collection of images for autonomous driving. The dataset includes different perspectives of the world captured at various intervals of time and the drivable range of vehicles, pedestrians from urban to residential, day and night, including sunrise, sunset, afternoon, morning, night and dawn. Both RGB and thermal cameras are coaligned to attain the same orientation in both images. A detailed examination of RGB, Thermal, a fusion of RGB and Thermal and LiDAR images along with the appropriate camera specifications is presented. Apart from object detection, this paper talks about visual perception tasks like drivable range detection, localization, image enhancement, depth estimation, and colorization using a single/multi-spectral approach and lessons about the progress of capturing data sets.[3]

This paper provides a proper definition for image fusion, stating that image fusion basically reads multiples images from the same scene, thus retrieving vital information to put into a single output image which is more informative and suitable for visual perception or computer processing, thereby improving the quality and applicability of the images. It also gives a brief literature review on various image fusion techniques in the spatial and frequency domain, such as averaging, min-max, block replace, HIS, PCA, brovey, pyramid-based techniques and, transform-based techniques. This paper also discusses various quality metrics for quantitative analysis of these approaches. [4]

This paper points out that the primary drawback of RGB semantic segmentation networks is that RGB images are susceptible to degradation with inadequate lighting conditions while thermal cameras generate images using thermal radiation emitted by objects and are independent of lighting conditions. So, this paper proposes to fuse both the RGB and thermal information in a novel deep neural network. An Encoder-Decoder design concept and ResNet is employed for feature extraction.[5]

This paper takes advantage of deep model-based methods of detection that have been effective in the visible domain and extends these approaches to the multispectral cases. This paper has made the first attempt to use deep models to combine the images of a visible and thermal camera. The introduced models and training methods have been evaluated on the KAIST multi-spectral pedestrian detection benchmark and have been compared to the state-of-the-art approaches. It shows that a pre-trained late-fusion based deep model outperforms the state-of-the-art approaches. [6]

This paper makes use of convolutional neural networks to generate the fusion of visible and thermal camera images to detect people present in those images, with a reliable surveillance system as an application in mind. It makes use of the Encoder-Decoder architecture for the fusion of visible and thermal imagery, additionally making use of the ResNet-152 architecture to detect the presence of the people in those images. The paper makes use of the KAIST multispectral dataset for training the CNNs, results show that fused architecture outperforms individual visible and thermal-based architectures. [7]

This paper compares six convolutional network fusion architectures and analyses their adaptations, enabling a vanilla architecture to obtain detection performances comparable to state-of-the-art results. Based on the discovery that pedestrian detection confidences from color to thermal images are correlated with illumination conditions, it proposes an illumination-aware Faster R-CNN to give an illumination measure of the input image. [8]

## References

1. Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi, “You Only Look Once: Uniﬁed, Real-Time Object Detection”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779 - 788, June 2016
2. Hongquan Qu, Tongyang Yuan\*, Zhiyong Sheng, Yuan Zhang North China University of Technology Beijing, China “A Pedestrian detection method Based on YOLOv3 model and Image enhanced by Retinex” 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2018)
3. A. González, Z. Fang, Y. Socarras, J. Serrat, D. Vázquez, Xu J., A.M. López, “Pedestrian detection at day/night time with visible and fir cameras: A comparison”, Sensors, 16 (6) (2016), p. 820
4. D. Mishra, B. Palkar, "Image Fusion Techniques: A Review", *International Journal of Computer Applications*, vol. 130, pp. 0975-8887, 2015.
5. Yuxiang Sun, Weixun Zuo and Ming Liu, “RTFNet: RGB-Thermal Fusion Network for Semantic Segmentation of Urban Scenes”*, IEEE Robotics and Automation Letters, Volume 4, July 2019, pp. 2576-2583*
6. J. Wagner, V. Fischer, M. Herman, and S. Behnke. Multispectral Pedestrian Detection using Deep Fusion Convolutional Neural Networks. In *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN),* 2016.
7. B. Khalid, A. M. Khan, M. U. Akram and S. Batool, "Person Detection by Fusion of Visible and Thermal Images Using Convolutional Neural Network," *2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE), Islamabad, Pakistan, 2019, pp. 143-148.*
8. C. Li, D. Song, R. Tong, M. Tang, “Illumination-aware Faster R-CNN for robust multispectral pedestrian detection,'' Pattern *Recognition, vol. 85C, (2019), pp. 161–171.*