https://doi.org/10.2352/ISSN.2470-1173.2019.8.IMAWM-414

Edge/Region Fusion Network for Scene Labeling in Infrared Imagery

Bradley Sorg; UD Vision Lab; University of Dayton; Dayton, OH; Theus Aspiras; UD Vision Lab; University of Dayton; Dayton, OH; Vijayan Asari; UD Vision Lab; University of Dayton; Dayton, OH

Abstract

Semantic segmentation has been a complex problem in the field of computer vision and is essential for image analysis tasks. Currently, most state-of-the-art algorithms rely on deep convolutional neural networks (DCNN) to perform this task. DCNNs are able to down-sample the spatial resolution of the input image into low resolution feature mappings which are then up-sampled to produce the segmented images. However, the reduction of this spatial information causes the high frequency details of the image to be lessened resulting in blurry and inaccurate object boundaries. In order to improve this limitation, we propose combining a DCNN used for semantic segmentation with semantic boundary information. This is done using a multi-task approach by incorporating a boundary detection network into the encoder decoder architecture SegNet. This multi-task approach includes the addition of an edge class to the SegNet architecture. In doing so, the multi-task learning network is provided more information, thus improving segmentation accuracy, specifically boundary delineation. This approach was tested on the RGB-NIR Scene dataset. Compared to using SegNet alone, we observe increased boundary segmentation accuracies using this approach. We are able to show that the addition of a boundary detection information significantly improves the semantic segmentation results of a DCNN.

Introduction

Semantic segmentation is one of the pivotal problems in the field of computer vision and machine learning. The task of semantic segmentation involves understanding an image at the pixel level by assigning each pixel in the image to an object class, which is essential for complete scene understanding. Many deep learning related tasks rely on this ability including robotic vision, autonomous driving, and medical imaging. Currently, most state of the art architectures used for semantic segmentation are deep convolutional neural networks.

Semantic segmentation deep learning architectures provide incredible results in segmentation and classification of various scenes, because of their ability to learn complete mappings from the raw input image to class labels. With their ability to capture the entire context of an image, these convolutional-based networks create deep representations for classification and have extended connected weight sets to improve the boundary characteristics of segmentations. However, this performance does comes at a cost. Because of the network's ability to cover a large context of the input image, the network must perform strong down-sampling to capture the entire image. This strong down-sampling results in the loss of high-frequency detail, particularly accurate localization of the object boundaries. To overcome this limitation, we propose a multi-task architecture for the deep learning encoder-decoder architecture SegNet to further improve the boundary characteristics of the neural network.

This paper aims to define a deep convolutional neural network with built-in edge/boundary detection to be used for semantic segmentation. Basic edge detection architectures are able to develop nice boundaries, but are unable to fully characterize the necessary boundary information from the imagery. To overcome this, we supplement the deep neural network architecture SegNet with specific boundary information to remove the boundary information that is not indicative of the boundaries of the classified regions. This is done with the addition of edge based learning to the architecture through the use of an edge class. By having the network learn to classify the relevant boundaries as a class, the network is biased towards identifying the correct and necessary boundaries of the different regions in the image.

To test and evaluate the multi-task SegNet architecture, we utilize the RGB-NIR Scene dataset for semantic segmentation. As compared to the original architecture, we observe an increase in boundary segmentation accuracy and boundary recreation using this approach. The incorporation of multi-task learning helps improve the semantic segmentation results of the deep learning architecture.

In this paper, we propose a multi-task SegNet architecture to further improve the boundary characteristics of the neural network. In this regard, the contributions of this paper can be summarized as follows:

- We introduce a multi-task approach for improved boundary delineation in a semantic segmentation architecture.
- We explore a multi-task combination of adding an edge class to the SegNet architecture.
- We show that a multi-task approach significantly improves the mean BF score in infrared imagery. In regard to the infrared imagery, this improvement comes with minimal difference in overall pixel accuracy.

The rest of the paper is structured as follows: After introducing semantic segmentation, explaining how it is applied in deep learning, and discussing multi-task learning in Chapter II, we give an overview of the SegNet architecture in Chapter III. In Chapter IV, we present our multi-task SegNet architecture for improved segmentations. We show the qualitative and quantitative training and testing performance of the network on the RGB-NIR Scene dataset and compare these performances against the original SegNet architecture in Chapter V. We then conclude and display potential future work in Chapter VI.

Background and related work

Semantic Segmentation

Semantic segmentation has become one of the key areas of interest and research in the field of computer vision and machine learning. Semantic segmentation involves understanding a given image at the pixel level. In doing so, every pixel of an image is able to be assigned to a given object class in that particular image.

In order to fully understand this concept, it is first important to comprehend what exactly segmentation is. Segmentation involves looking at a digital image and partitioning it into different segments.

Typically, an image is inputted into a system, the system processes the image, then outputs a different representation of this image. Normally, this image is changed into something that is more meaningful and easier to analyze. In most cases, segmentation is used to identify different objects or boundaries in a given image. This technique differs from object recognition or image classification in the sense that it is not necessary for the system to know what objects to look for in the images beforehand. For example, an object classification algorithm will only be able to classify objects that it has a specific label for, like a person, dog, tree, or building. An ideal semantic segmentation algorithm, however, will be able to segment unknown objects in the image which are new or unknown to the system beforehand. For instance, when given a new image, an image segmentation algorithm should be able to output which pixels in the image semantically belong together. Unlike classification algorithms, we need these models to be able to make dense pixel-wise predictions.

For a human, performing this task would seem trivial. However, for a computer this task is much more difficult making semantic segmentation a current area of interest for researchers and engineers. Semantic segmentation is a difficult task as a result of the computer needing to have a deep understanding of the image. Compared to other computer vision tasks like object recognition or image classification, semantic segmentation involves an understanding of an image at the pixel-level. Because of this, the details and features of the image need to be understood by the system in much greater detail than other deep learning tasks.

Deep Learning Applications

Because of its ability to assist in complete scene understanding, semantic segmentation has become an important area in the field of computer vision and machine learning. Since most of the applications in these fields thrive on being able to infer knowledge from imagery, semantic segmentation is a very useful technology and tool. Three of the most promising and researched applications are robotic vision, autonomous driving, and medical imaging.

In the field of robotics, being able to segment the surrounding environment for a robot is a very important task. This would provide the robot with information about how the environment around it is structured as well as give information about objects that the robot may come in contact with. Meyer and Drummond [1] have introduced a technique using semantic segmentation that results in significantly less false positive object detections for robotic vision applications. Wolf et al. [2] introduced a three-dimensional (3-D) entangled forests technique to improve the semantic segmentation of 3-D point clouds. This technique is able to learn and exploit common contextual relations between observed structures and objects which is essential for robotic vision tasks. As segmentation algorithms continue to improve, robotic vision applications will only become more and more advanced.

Similar to robotic vision, autonomous driving is another application that benefits from image segmentation. Image segmentation aids in being able to understand the vehicle's surroundings, like its spatial relationship with the other cars on the road or other objects in the scene. By equipping cars with this necessary perception, self-driving cars will be able to be safely integrated onto the road.

This technique is even being used to model the different traffic patterns a car may experience. In Zhang and Geiger [3], a generative model of 3D urban scenes is produced using image segmentation. This model produced an improved overall scene estimation by assisting in associating objects with the correct lanes. While most of the current segmentation algorithms are designed for generic images, they are beginning to incorporate prior structure in the data to aid in the autonomous driving problem.

The field of medical imaging is another area that benefits greatly from advancements in semantic segmentation. Health care providers rely on medical images in order to correctly assess patients for diagnosis and treatment. More often than not, the studying of these medical images is performed by a radiologist. As a result, the studying of these images is a direct result of the visual interpretation by the radiologist. This process is normally time consuming and subjective, as the visual interpretation is often a result of the experience of the radiologist. To overcome these limitations, computer systems that implement segmentation are beginning to be introduced in the medical field. For example, Narayanan, Hardie, and Kebede [4] are able to show that a computer-aided detection system is more successful in detecting lung nodules than the typical CT scans and chest radiographs radiologists use to detect lung nodules. Gooya et al. [5] designed a software package, GLISTR, aimed at simultaneously segmenting brain scans of glioma patients. They were then able to use these scans to construct a statistical atlas of the glioma, a type of tumor. Improvements made in medical image segmentation will lead to more accurate diagnoses as well as better care and treatment for patients.

There are numerous other applications that image segmentation can be applied to as well. Chatfield and Arandjelovic [6] explore its use in on-the-fly visual search in addition to content-based image retrieval. For example, an image could be segmented and added to a database. This would allow a user to query and look for particular entities in the database, like all images containing an airplane. Yu and Qin [7] discuss how this technique could be applied to SAR image processing for better visual representations. With all of the applications mentioned, having access to segmentations would allow these problems to be approached at a semantic level.

Multi-Task Learning

Multi-task learning is a subfield of machine learning that has become popular as of late. Multi-task learning is described as learning multiple tasks at the same time while utilizing the commonalities and differences between the tasks. In doing so, this can result in improved learning efficiency and prediction accuracy for the multi-task model, compared to training the models independently. As far as its application in machine learning, multi-task learning has aided in speech recognition [8], drug discovery [9], food category/calorie estimation [10], natural language processing [11], medical diagnoses [12], and computer vision [13] related tasks.

Related to the task of semantic segmentation, multi-task learning has become an area of interest due to its ability to improve prediction accuracies. Currently, most research regarding using multi-task learning for semantic segmentation pertains to the incorporation of boundary information into a neural network architecture. Most state of the art semantic segmentation architectures perform exceptionally well in their ability to classify pixels with their correct class, but often lack in their ability to delineate boundaries between the different classes. As a result, by making the class boundary information available to the network, the network should be able to produce more accurate and distinct segmentations.

Recently, researchers have had success in improving segmentation results by incorporating boundary information. Marmanis et al. [14] work to preserve boundary information on segmentation classes by incorporating an edge detection network into a neural network. Using SegNet [15] as a feature extractor and the holistically-nested edge detection (HED) network [16,17] to detect edge information, the boundary information is fed into the network by concatenating the feature maps of SegNet with the edge prediction. Using this approach, Marmanis et al. [14] saw an improvement of

labeling accuracy of up to 6% for the ISPRS semantic labeling benchmark.

Bischke et al. [18] employed a similar approach for the semantic segmentation of high resolution satellite imagery. Their cascaded multi-task network is a single network containing a single loss function comprised of two pixel-wise classification losses, one for semantic information and the other for geometric properties. This approach was able to successfully increase the accuracy for the segmentation of building footprints in remote sensing imagery. Mou and Zhu [19] were able to use a modified version of the residual neural network ResNet to construct a fully convolutional network used for vehicle instance segmentation from aerial imagery and video. Differing from semantic segmentation, instance segmentation involves identifying, at a pixel-level, where the vehicle appears as well as associating each pixel with a physical instance of a vehicle. As can be seen, incorporating boundary information through multi-task learning has had a positive impact on improving semantic segmentation results.

Overview of SegNet

SegNet, introduced in [15], is a variation of a convolutional neural network (CNN). SegNet is an end to end convolutional encoder decoder neural network architecture used for semantic segmentation. SegNet, as illutstrated in Figure 1, includes an encoder network with a corresponding decoder network followed by a final pixelwise classification layer.

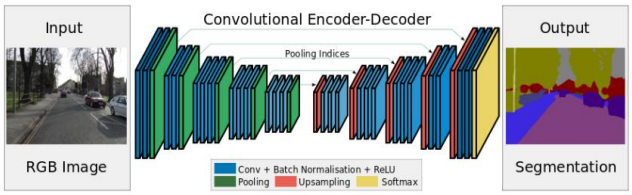


Figure 1: An illustration of the SegNet architecture [15].

The encoder network is made up of 13 convolutional layers which are identical to the first 13 convolutional layers found in the VGG16 network. By using this type of encoder network, SegNet is able to omit fully connected layers. In turn, higher resolution feature maps are able to be retained deeper into the encoder network. This also allows the encoder network to significantly reduce the number of parameters used compared to other architectures found in [20] and [21]. The decoder network is also made up of 13 layers as each encoder layer has a corresponding decoder layer. The output of the final decoder layer is then fed into a multi-class soft-max classifier that produces the class probabilities for each pixel in an image.

Looking deeper at the encoder network, each encoder layer performs a convolution with a filter bank to produce a set of feature maps. These feature maps are then batch normalized [22] and applied to a rectified non-linearity (ReLU) [23] of $max(0,x)$, as seen below in Equation 1

$$f(x) = x^{+} = \max(0, x). \quad (1)$$

In Equation 1 above, the ReLU is applied element-wise where it produces an output of x if x is positive, otherwise it outputs a 0. Following this two-step process, max-pooling is performed using a 2x2 window with a stride of 2. This is used as a way to achieve translation invariance over small spatial shifts in the input image. This reduces the spatial size representation of the image which in turn reduces the amount of parameters and computations in the network resulting in more efficient computations. By sub-sampling the input image, a larger context of the image is able to be covered for each pixel in the feature map.

Although several layers of max-pooling and sub-sampling can achieve more translation invariance, this approach results in significant loss of spatial resolution, particularly boundary detail, of the feature maps. As a result, the boundary information in the image needs to be captured and stored in the encoder network's feature maps before any sub-sampling is performed. To overcome this limitation, SegNet reuses the max-pooling indices, the locations of the maximum feature value in each pooling window, from each encoder feature map. This makes SegNet much more memory efficient since only the max-pooling indices are copied as opposed to all of the encoder features.

SegNet is further able to make use of max-pooling indices in the decoder portion of the network. Each decoder of the decoder network of the architecture is able to use the max-pooling indices from its corresponding encoder feature map to up-sample the input feature maps. This results is sparse feature maps being produced. Figure 2 shows an illustration of the decoding process used.

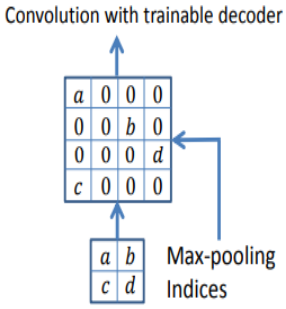


Figure 2: An illustration of the SegNet decoder functionality.

In understanding the decoding process above, *a, b, c, d* correspond to values in a feature map. As can be seen, SegNet is able to reuse the max pooling indices to up-sample the feature map without learning the entire feature map itself. This differs from other architectures, particularly U-Net, where the entire feature map is transferred to the corresponding decoder, resulting in more memory being used. These up-sampled feature maps are then convolved with a trainable decoder filter bank to produce dense feature maps. The dense feature maps are in turn batch normalized. This process is repeated until the final decoder of the network is reached. The output of the final decoder is then fed to a trainable soft-max classifier which is able to classify each pixel independently. The soft-max classifier outputs a K channel image of probabilities where K corresponds to the number of classes present in the image. Each pixel is then assigned the class with the maximum probability resulting in the predicted segmentation. Thus, a segmented image representing the predicted class labels is produced.

Although SegNet reuses the max-pooling indices, spatial information of the feature maps is still lost. Most semantic segmentation architectures suffer from this problem, thus the boundary details of the segmented outputs are often blurry and non-delineated. Because of the memory efficient nature of SegNet's architecture, a multi-task approach is explored in Chapter IV to incorporate boundary detection awareness into this network design.

Multi-Task SegNet Architecture for Semantic Segmentation

The multi-task approach to incorporate additional boundary information from the imagery into the SegNet architecture was to introduce an edge class label to the network. This is done by detecting the edges in the mask containing the class labels for a given image in

a dataset. In order to do this, the Prewitt operator is used. Developed by Judith M.S. Prewitt [24], the Prewitt operator calculates the gradient of the image intensity at each point in an image. The Prewitt operator gives the direction and magnitude of the largest possible pixel increase from light to dark in an image. As a result, it is able to detect how "abruptly" or "smoothly" the image changes at that point determining whether or not that part of the image represents an edge.

From a mathematical perspective, the Prewitt operator uses two 3x3 kernels that are convolved with the original image to calculate approximations of the horizontal and vertical changes in the form of two derivatives. Equation 2 shows this process below

$$\boldsymbol{G_X} = \begin{bmatrix} -1 & 0 & +1 \\ -1 & 0 & +1 \\ -1 & 0 & +1 \end{bmatrix} * \boldsymbol{A} \; and \; \boldsymbol{G_X} = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * \boldsymbol{A} \tag{2}$$

where **A** is the source image, $\mathbf{G_x}$ and $\mathbf{G_y}$ are two images that contain the horizontal and vertical derivative approximations for each point in the image, and * denotes the 2-dimensional convolution operation. The gradient approximation at each point in the image can then be calculated using

$$\boldsymbol{G} = \sqrt{\boldsymbol{G_x^2} + \boldsymbol{G_y^2}} \tag{3}$$

resulting in the gradient magnitude. Applying the Prewitt operator on the label mask for a given image results in a matrix. This matrix is the same size as the label mask and contains the gradient magnitude for each pixel. By taking the magnitude values that are greater than zero, the pixel locations for the edges or boundaries between the different regions in the image can be identified. These pixel locations are then updated in the original label mask and are given the label associated with the edge class. By including this edge class, the network is now biased towards learning the edge/boundary information for all images in a dataset.

Results and discussion

To evaluate the performance of our multi-task SegNet architecture compared to the regular SegNet architecture, the RGB-NIR Scene dataset was used. For each architecture, the global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BF score metrics were calculated for the training and testing sets for the dataset. Both of these neural network architectures were implemented in MATLAB and ran on two GeForce GTX 1080 Ti's GPUs. In order to perform a controlled training procedure, each architecture was trained with an initial learning rate of 10^{-3} using 10 epochs and a batch size of 20. The learning rate was then lowered to 10^{-6} and each architecture was again trained using the same number of epochs and batch size.

RGB-NIR Scene Dataset

To evaluate the multi-task approach on infrared imagery, the dataset used for evaluation was the RGB-NIR Scene Dataset [25]. This dataset consists of 477 images captured in RGB and near-infrared (NIR). The images were captured using separate exposures from modified SLR cameras, using visible and NIR filters. This dataset did not include any pixel-level labeling, so pixel labeling was done at our discretion. The dataset is broken into two categories: indoor and outdoor scenes. To evaluate these architectures on infrared imagery, the NIR images from the outdoor scenes were used. The outdoor scenes are comprised of 370 images, so training and testing sets were created using 185 images for each. These images were scaled into smaller sub-images, roughly 120,000 in each set, in order to cover the entire context of each image. Using these sub-images, the initial SegNet architecture, as well as the multi-task SegNet architecture, were trained and tested. Figures 3 and 4 show a qualitative comparison of the predictions made by the different architectures for the training and testing sets.

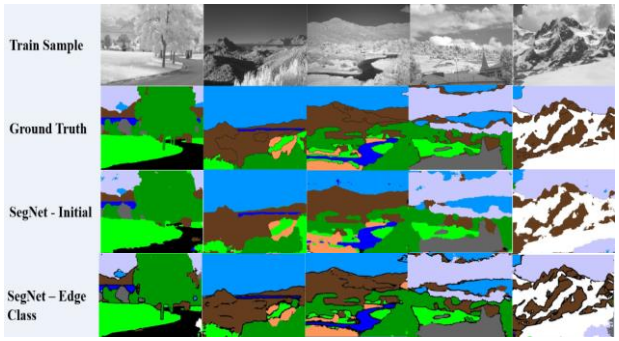


Figure 3: Qualitative results for the training set of the RGB-NIR Scene dataset.

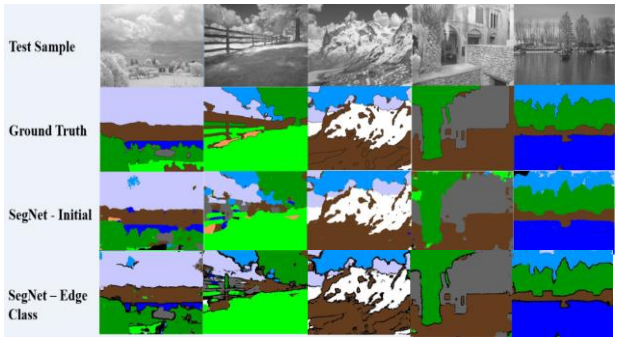


Figure 4: Qualitative results for the testing set of the RGB-NIR Scene dataset.

In comparing the qualitative results between the two architectures, there are a few key observations. Comparing the initial SegNet architecture to the SegNet edge class architecture, the edge class architecture seems to delineate the correct boundary information much better than the initial architecture. The edge class architecture produces much more distinct boundaries, resulting in much clearer segmentations. Even if the classifications are wrong, the more distinct boundaries make it much easier to visualize what the object in the image should be. For example, in the second testing sample, seen below in Figure 5, although the fence is misclassified in both the initial and edge class architectures, it is much easier to see that the object is a fence in the edge class produced image.

Figure 5: Second testing sample found in Figure V.2 for the RGB-NIR Scene dataset.

The qualitative results show the ability of the proposed edge class SegNet architecture to segment smaller classes in infrared imagery while also producing a smooth segmentation of the overall scene.

To further compare their performances, a quantitative comparison is made using the global accuracy, mean accuracy, mean IoU, weighted IoU, and mean BF score metrics. Tables 1 and 2 show a quantitative comparison of the different architectures for the training and testing sets of the modified RGB-NIR Scene Dataset.

Table 1: The metric values calculated for each architecture using the training set for the RGB-NIR Scene dataset.

Architecture	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	Mean BFScore
SegNet	0.95257	0.93882	0.89621	0.91033	0.62212
SegNet with Edge Class	0.94547	0.92579	0.87974	0.90595	0.73364

Table 2: The metric values calculated for each architecture using the testing set for the RGB-NIR Scene dataset.

Architecture	Global Accuracy	Mean Accuracy	Mean IoU	Weighted IoU	MeanBF Score
SegNet	0.65675	0.60366	0.45782	0.49315	0.22018
SegNet with Edge Class	0.63464	0.58162	0.43863	0.47896	0.36506

In comparing the quantitative results between the two architectures, it can be seen that the qualitative observations are validated. The initial SegNet architecture outperformed the multi-task approach in terms of global accuracy, however only slightly. The mean BF score of the edge class architecture is significantly higher than the original SegNet architecture with an 11% increase on the training set and a 14% increase on the testing set. The global accuracy difference between the initial and edge class SegNet architectures is minimal. The loss in global accuracy for the edge class architecture can be attributed to the removing of class labels when adding the edge class. However, since there are a less number of classes found in this dataset, the removal of these labels doesn't seem to have as great an effect on the overall global accuracy.

In terms of comparing the training and testing results for both architectures, the overall decrease in performance metrics for the testing data can be attributed to the variation found in the dataset. The RGB-NIR Scene dataset has a lot of variation as there isn't much commonality between images. As a result, the testing set introduces a lot of variability to the networks in terms of imagery the network has never seen before.

For the improvement seen in the mean BF score, roughly 11% and 14% increases on the training and testing sets, the edge class architecture is superior to the initial SegNet architecture. The minimal difference in global accuracy, less than 1% on the training set and 3% for the testing set, is a justifiable trade-off for such a significant improvement in boundary delineation accuracy. As a result, the multi-task edge class approach results in improved boundary segmentations compared to the initial SegNet architecture.

Conclusion and Future Work

In this paper, it was seen that the incorporation of learned boundary information significantly improves the boundary characteristics of a neural network. The multi-task approach of incorporating an edge class into the SegNet architecture outperformed the initial SegNet architecture in terms of mean BF score on the RGB-NIR Scene dataset. Using this approach, improved boundary segmentation accuracies and boundary recreations were observed. By having the network learn to classify the relevant boundaries as a class, the network was biased towards identifying the correct and necessary boundaries of the different regions in the image.

For future work, other methods are being researched to incorporate additional boundary information into neural networks. Other multi-task learning approaches can be developed to assist in this task. The successful approach of having the network learn the boundaries as a class could be extended to having the network serve as an edge detector. If a given network is simply trained as an edge detector, the network will then be able to identify the correct boundary information in an image. This output could then be combined with the segmented output to produce clearer segmentations. Overall, improvements to the boundary characteristics of a neural network is still desired.

References

[1] B. J. Meyer and T. Drummond, "Improved semantic segmentation for robotic applications with hierarchical conditional random fields," 2017 *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, pp. 5258-5265, 2017.

[2] Wolf, D., Prankl, J., Vincze, M., "Enhancing Semantic Segmentations for Robotics: The Power of 3-D Entangled Forests," *IEEE Robotics and Automation Letters,* vol, 1, pp. 49-56, Jan. 2016.

[3] H. Zhang, A. Geiger, and R. Urtasun, "Understanding high-level semantics by modeling traffic patterns," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3056–3063.

[4] B.N. Narayanan, R.C. Hardie, and T.M. Kebede, "Performance Analysis of a Computer Aided Detection System for Lung Nodules in CT at Different Slice Thicknesses", *SPIE Journal of Medical Imaging,* 5(1) 014504(2018). Doi:10.1117/1.JMI.5.1.014504.

[5] A. Gooya, K.M. Pohl, M. Bilello, L. Cirillo, G. Biros, E.R. Melhem,and C. Davatzikos, "GLISTR: Glioma Image Segmentation and Registration," *IEEE Trans. Med. Imaging,* 31(10): 1941-1954 (2012).

[6] K. Chatfield, R. Arandejelovic, O. Parkhi, and A. Zisserman, "On-the-fly learning for visual search of large-scale image and video datasets," *International Journal of Multimedia Information Retrieval,* 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4498639/.

[7] P. Yu, A.K. Qin, and D.A. Clausi, " Unsupervised Polarimetric SAR Image Segmentation and Classification Using Region Growing With Edge Penalty," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 50. pp. 1302-1317. Sept. 2011.

[8] L. Deng, G.E. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: An overview," *IEEE International Conference on Acoustics, Speech, and Signal Processing,* pp. 8599-8603, 2013.

[9] B. Ramsundar, S. Kearnes, P. Riley, D. Webster, D. Konerding, and V. Pande, "Massively Multitask Networks for Drug Discovery," Feb. 2015. [Online]. Available: https://arxiv.org/abs/1502.02072.

[10] T. Ege and K Yanai. "Simultaneous Estimation of Food Categories and Calories with Multi-task CNN," in *Proc. of IAPR International Conference on Machine Vision Applications (MVA)*, 2017.

[11] R. Collobert, and J. Weston, "A unified architecture for natural language processing," *Proceedings of the 25th International Conference on Machine Learning,* vol. 20. pp. 160-167, 2008.

[12] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease," *NeuroImage,* vol. 59. pp. 895-907, 2012.

[13] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1440-1448, 2015.

[14] D. Marmanis, K. Schindler, J. D. Wegner, S. Galliani, M. Datcu, and U. Stilla, "Classification with an edge: Improving semantic image segmentation with boundary detection," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 135, pp. 158-172, Jan. 2018.

[15] V. Badrinarayanan, A. Kendall and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481-2495, 1 Dec. 2017.

[16] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1395–1403, 2015.

[17] I. Kokkinos, "Pushing the boundaries of boundary detection using deep learning," in *International Conference on Learning Representations,* 2016.

[18] B. Bischke, P. Helber, J. Folz, D. Borth and A. Dengel, "Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 2018.

[19] L. Mou and X.X. Zhu, "Vehicle instance segmentation from aerial image and video using a multitask learning residual fully convolutional 14 network," in *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2018.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, 2015.

[21] N. Hyeonwoo, H. Seunghoon, H. Bohyung, "Learning Deconvolution Network for Semantic Segmentation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV),* pp. 1520-1528, 2015.

[22] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *CORR*, vol. abs/1502.03167, 2015.

[23] F. Abien and M. Agarap, "Deep Learning using Rectified Linear Unites (ReLU)," *Neural and Evolutionary Computing,* vol. 1, 2018.

[24] J. M. S. Prewitt, "Object enhancement and extraction," *Picture Processing and Psychopictorics*, New York: Academic Press, pp. 75-149, 1970.

[25] M. Brown and S. Süsstrunk, "Multispectral SIFT for Scene Category Recognition," *International Conference on Computer Vision and Pattern Recognition (CVPR),* 2011.

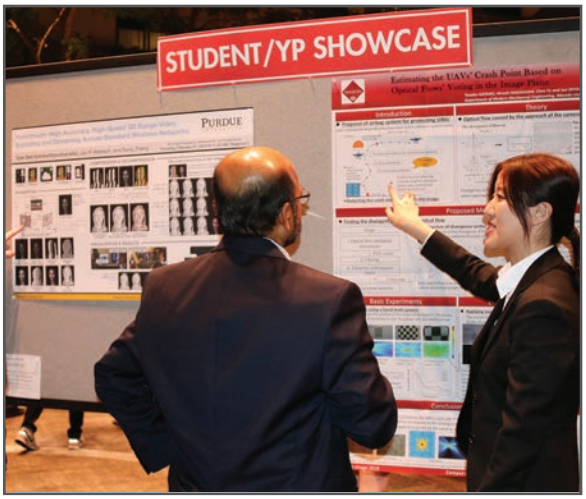
JOIN US AT THE NEXT EI!

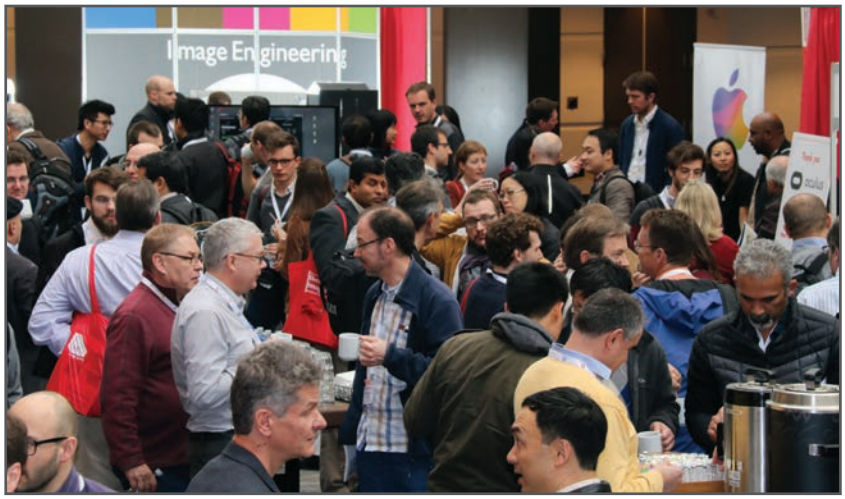
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org